

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 9, 2004, 21:06:31 ; Search time 4943 Seconds
(without alignments)
4682.013 Million cell updates/sec

Title: US-10-726-148a-15
Perfect score: 4712
Sequence: 1 MLQQLPVSLLSVAVSAIK.....PFTVKNPMSPTPKWNPKA 775

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.3
Delop 6.0 , Delext 7.3

Searched: 27513289 seqs, 14931030276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2 1/USPTO.spool/J510726148/runat 09062004_61329 5290/app query.fasta_1.967
-DB=EST -OPWT=fastap -SUFFIX=p2n.rst -MINMATCH=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10726148.scgn 1 1 3609/runat 09062004 161329 5290 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estmu:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_nam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	816	17.3	1116	12	BM459284	BM459284 AGENCOURT
2	802	17.0	955	13	BQ714462	BQ714462 AGENCOURT
3	674.5	14.3	3036	11	AK089086	AK089086 Mus muscu
4	668	14.2	2277	11	AK089169	AK089169 Mus muscu
5	665.5	14.1	744	10	BE538260	BE538260 601062138
6	589.5	12.5	1595	11	AK052479	AK052479 Mus muscu
7	578.5	12.3	3713	11	AK036800	AK036800 Mus muscu
c	577	12.2	697	12	BQ045321	BQ045321 UI-CF-EN1
8	563.5	12.0	667	10	BB629805	BB629805 BB629805
9	558	11.8	877	12	BM560443	BM560443 AGENCOURT
10	545.5	11.6	3108	11	AF328918	AF328918 Homo sapi
11	540	11.5	634	10	BB660137	BB660137 BB660137
12	522.5	11.1	2453	11	BC026542	BC026542 Mus muscu
13	511.5	10.9	675	13	BY721112	BY721112 BY721112
14	496	10.5	357	12	BQ028158	BQ028158 UI-H-COO-
c	488	10.4	510	10	BF833247	BF833247 PM3-HT090
16	488	10.4	510	10	BF833459	BF833459 PM3-HT090
c	473.5	10.1	2661	29	AY399360	AY399360 Homo sapi
19	474.5	10.1	3782	11	AK041860	AK041860 Mus muscu
20	468.5	9.9	915	29	CG947840	CG947840 MBEME30TR
c	451.5	9.6	2607	29	AY399361	AY399361 Pan trogl
23	446	9.5	521	10	BF062477	BF062477 7h59g03.X
24	445	9.4	636	13	BQ687111	BQ687111 UI-CF-DUI
25	439	9.3	2664	29	AY399362	AY399362 Mus muscu
26	433.5	9.2	862	13	BU089166	BU089166 AGENCOURT
27	431	9.1	2376	11	AK048901	AK048901 Mus muscu
28	429.5	9.1	2675	11	BC051074	BC051074 Mus muscu
29	427	9.1	2531	11	AK039384	AK039384 Mus muscu
30	421	8.9	3881	11	BC047156	BC047156 Mus muscu
31	418.5	8.9	3114	11	BC034057	BC034057 Mus muscu
c	396	8.4	348	10	AW879171	AW879171 MRI-OT001
c	395.5	8.4	338	14	CK270886	CK270886 EST716964
33	395.5	8.4	4782	11	AK048546	AK048546 Mus muscu
34	395.5	8.4	3751	11	AK080277	AK080277 Mus muscu
35	394	8.4	3751	11	AK080277	AK080277 Mus muscu
36	391.5	8.3	4784	11	AK041115	AK041115 Mus muscu
37	387	8.2	837	12	BI756263	BI756263 603024322
c	386.5	8.2	1049	29	CNS031TO	AL249909 Tetraodon
c	386.5	8.2	1125	29	CNS057N6	AL324843 Tetraodon
40	384.5	8.2	396	9	AA280392	AA280392 zt04h04.r
41	383	8.1	450	10	AM605922	AM605922 RCL-HT025
42	383	8.1	3577	11	AK053825	AK053825 Mus muscu
43	380	8.1	4783	11	AK029212	AK029212 Mus muscu
44	377.5	8.0	780	13	BX482979	BX482979 DKFZp686A
c	370	7.9	732	14	CK268212	CK268212 EST714290

ALIGNMENTS

RESULT 1
BM459284
LOCUS BM459284 1116 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6415560 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5495835
5', mRNA sequence.
ACCESSION BM459284.1 GI:18508324
VERSION BM459284
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1116)

Db	162	ACCAAGACCCAGACGCAAGCAACA-TTGAACCTGAGCTAAGCTACAAATGACAGTA	221
Qy	64	AsnGlyIysIleAlaValLeuTyrIleuLysIysAsnLysAsnLeuLeuAlaProGlyTyr	83
Db	222	AATGGAAGGTTGCTGTCTGTATCTGAAGAAGAACAAACAGCTCTTTCGCGCTGACTAC	281
Qy	84	IhrGluThrCyrTyrAsnSerThrGlyLysGluIleThrThrSerProGlnIleMetTrp	103
Db	282	TCGGANAACATCTAATTCAGTCCAGTCAACAGGTCACCAAGCCGCCAAATCATGAT	341
Qy	104	TrpTrpTyrGlnGlyHisIleLeuAsnGluLysValSerTrpAlaSerIleSerThr	123
Db	342	AGCTGTTACTACCAAGGACACATCTGAATGAGAAAGTTCTCGACCCAGCATCAGCAC	401
Qy	124	TrpArgGlyLeuArgGlyTyrTrpSerGlnGlyTrpGlnArgTyrTrpIleGluProLeu	143
Db	402	TSTCAAGGACTACGCGGTTTACATCATCAAGGAGATGAAGATATTTATCATCAACCTTG	461
Qy	144	SerProIleHisArgTrpGlyGlnGluHisAlaLeuTyrLysTyrAsnProTrpGluLys	163
Db	462	AGCTCGAGACCTGGATGAACAGGCACATGCATCTTCAAGGACGACTCCATGAAGAC	521
Qy	164	AsnTyrTrpSerThrTrpGlyMetTrpGlyValLeuTyrAlaHisTrpLeuGlnAsn	183
Db	522	CAGGAAAGACCACTGTGTGGATGATGGCTATGGCTCAAGGGCTGCATCAGGAC	581
Qy	184	IleAlaLeuProAlaThrIleuValLysLeuLysTrpArgLysValGlnHisGlu	203
Db	582	GTGGCCCTCCCTGCCACCAAGTGAATTAAGTTGAATGATGGATGGTTCAGAACTAAG	641
Qy	204	LysTyrIleGluTyrTrpLeuValLeuTyrAsnGlyGluTyrLysArgTyrAsnGluAsn	223
Db	642	ANATACATAGATATTTGTGCTCTGGATATATGTTGAGTTTGAATAATACATAAAT	701
Qy	224	GlnTrpGluIleArgLysArgValTrpGluMetAlaAsnTyrValAsnMetLeuTyrLys	243
Db	702	CTTGCTGAATACGAAAGATAGTGTTCGATGGCCCAATTACATCAACATGCTTTACAT	761
Qy	244	LysLeuAsnThrHis-ValAla-LeuValGlyMetGlu-IleTrp-ThrTrpLysTrpLys	262
Db	762	ARGCTTGATGGCCACCGTGGCCCTTAGTTGGAGTGAATAATCTGNNACGATGGGGATAA	821
Qy	262	SileLysIleThrProAsnAlaSerTrpTrpLeuGlu-AsnTrpSerLysTrpArgGly	281
Db	822	ATAAAGATAACCCCGAGCCACACACACCCCTGGAAAATCTCTAGTGGAGAGGGA	881
Qy	282	SerValLeuSerArgArgLysArg-----HisTrpIleAlaGlnLeuIleThrAlaThr	299
Db	882	ATGATGCTGCTAAACCAAGCATCATGATATGG---CCCAGCCTATTCTCATCAACA	938
Qy	300	Glu 300	
Db	939	GAC 941	
RESULT	3		
LOCUS	AK089086	3036 bp	mx3c linear HTC 20-SEP-2003
DEFINITION	Mus musculus 2 days neonate thymus thymic cells cDNA, RIKEN full-length enriched library, clone:5430039A18 product:a disintegrin and metalloprotease domain 8, full insert sequence.		
ACCESSION	AK089086		
VERSION	AK089086.1	GI:26354229	
KEYWORDS	HTC; CAP trapper		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
1			
AUTHORS	Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE			
AUTHORS	Carninci, P., Hayashizaki, Y., Sugahara, Y., Sugahara, Y., Carninci, P., Konno, H., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20493374		
PUBMED	11042159		
REFERENCE			
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsumi, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Hayashi, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujikake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, C., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		
MEDLINE	20530913		
PUBMED	11076861		
REFERENCE			
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.		
TITLE	Functional annotation of a full-length mouse cDNA collection		
JOURNAL	Nature 409, 685-690 (2001)		
REFERENCE			
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.		
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs		
JOURNAL	Nature 420, 563-573 (2002)		
REFERENCE			
AUTHORS	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)		
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/WRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge. Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/ Location/Qualifiers		
FEATURES	1..3036		
source	/organism="Mus musculus"		
	/mol_type="mRNA"		
	/strain="M0D"		
	/db_xref="FANTOM DB:E430039A18"		
	/db_xref="MGI:2427983"		
	/db_xref="taxon:10090"		

Db 1506 CTACTGCTTTGATGGAGCTCTCCACCCCTG3GCACAGCGTGGCGGAGTCTGTGGGGGCC 1565

Qy 522 oGlyThrGluValAlaTrpLysSerTrpTrpAsnArgAsnGluGlyGlySerLysTrpG1 542
|||||
Db 1566 AGGTGCTGGGTAGCAGCGGACTCTGCTATACCTTTAGC----- 1605

Qy 542 yTyTrpArgArgValTrpTrpThrLeuIleProTrpLysAlaAsnTrpThrMetTrp-- 561
|||||
Db 1606 -----ATCCCTCCGGGCTGCAATGGGAGGAGTGTACTC 1637

Qy 562 -----GlyLysLeuTrpTrpGlnGlyGlySerTrpAsnLeuProTrp 575
|||||
Db 1638 TGGCAGAGTCAACCGGTGGAGGCTGTACTGTGAGGAGGC-----CAGAGCCCTC 1691

Qy 575 pLysGlyArgIleValTrpTrpLeuThrTrpLysThrTrpTrpProGluTrpThrSerG1 595
|||||
Db 1692 TGAACGCTCTCTGCACTTCTCTCC-----AACCATGGAGTGGCCA 1736

Qy 595 nGluIleGly-----MetValAlaAsnGlyThrLysTrp 606
|||||
Db 1737 TGCTCTTGGCACAGGCAACATTCGACCTTTGAGCTGGTATGTCAGGCGCAAGTGTG 1796

Qy 606 pGlyTrpAsnLysValTrpIleAsnAlaGluTrpValTrpIleGluLysAlaTrpLysSe 626
|||||
Db 1797 CGAGGAGGAAGGTTTCATGGATGGAGCTGCAGGACCTC---CGTGTATACAGATC 1853

Qy 626 rThrAsnTrpSerSerLysTrpLysGlyHisAlaValTrpTrpHisGluLeuGlnTrpG1 646
|||||
Db 1854 TGAATACTGCTCTCTAAATCAACACCATGGGTATGCACCAAGAGGAGTGGCCA 1913

Qy 646 nTrpGluGluGlyTrpIlePro-Pro-----TrpTrpTrpTrp----- 658
|||||
Db 1914 CTGTCAACAGGGCTGGGCACCACTGTGTACAGCGGCTGGCAGATGTACAGATGA 1973

Qy 659 -----SerSerValValTrpHisTrpSerIleValValGlyValL 672
|||||
Db 1974 ACAAGCAGCGTCTAGCAGCTCCAGTCCAGTGGTGGTGGTCTG-CTGATCTGGTG- 2031

Qy 672 euTrpProMetAlaValIleTrpValValValAlaMetValIleArgHisGlnSerSerA 692
|||||
Db 2032 -----GCTCGCATGTCATCGTGGCAGGCATCGTCATCTACCAAGGCTCGCA 2080

Qy 692 rgGluLysGlnLysLysTrpGlnArgProLeuSerThrThrGly---ThrArgProHisL 711
|||||
Db 2081 GACAAA-CCACAGGAGGAGTGTGGCAGCCCAAGCCCTATCTCGGGGCTCTCCAAACCCCTAT 2140

Qy 711 ySglnLysArgLysProGlnMetValLysA-avalGlnProGlnGluMetSerGlnMetL 731
|||||
Db 2141 TCTACACAGGAGCAGCAGCGCTGGCAGCTAAGACAGCGCTCCAGACCTTCTGAG--- 2196

Qy 731 ySProHisValTyTrpLeuProValGluGlyAsnGluProProAlaSer-----TrpH 749
|||||
Db 2197 -----ACAGTTTCTACCAACAGCCGCCCAAGACCCCATAGTGAAC 2236

Qy 749 iSlysTrpThrAsnAlaLeuProProThrValTrpLysTrpAsnProMetSerThrPro 769
|||||
Db 2237 CAAGAGCCCTCCCGCTCCACTCCAGTCTGTGTCCAGTTCACCACTCCAGTTCCTG 2296

Qy 769 ySTrpSerAsnProLys 774
|||||
Db 2297 TTTATGCC---CCAAAG 2310

RESULT 4
AK089169
LOCUS
DEFINITION Mus musculus MOD-derived CD11c +ve dendritic cells cDNA, RIKEN full-length enriched library, clone:F630003B03 product:a disintegrin and metalloprotease domain 8, full insert sequence.
AK089169
ACCESSION
VERSION AK089169.1 GI:26354293
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636

2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159

3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Smit, N., Ishii, Y., Nakamura, S., Hazama, M., Kishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, S., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

4
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
11076861

5
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2277)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, K., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.

COMMENT

FEATURES

Location/Qualifiers
 1. .2277
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="NOD"
 /db_xref="FANTOM_DB:F630003B03"
 /db_xref="MGI:2428075"
 /db_xref="taxon:10090"
 /clone="F63003B03"
 /cell_type="NOD-derived CD11c +ve dendritic cells"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 44. .2275
 /note="unnamed protein product; a disintegrin and metalloprotease domain 8 (MGD|MGI:107825, GS|NM_007403, evidence: BLASTN, 99%, match=3030)
 putative"
 /codon_start=1
 /protein_id="BAC40775.1"
 /db_xref="GI:26354294"
 /translation="MLGLWLSVLTTPVAFPPPLPHVKOVYEVVWPRLAASRRALP
 SHWQTPESLSYALG-SGVFTLHLKRNRLDLSSTETYSIAANGSEVTEQLQEDHC
 LYGHVEYEGSAS-STCAGLGRFPRVGSIVHLIELPDADEGQAMQAKHQKA
 CTGVKTNLDIGPALLEIYRAQPNWLPVRETVLYVADSOEFQKLGREAVR
 QVLEVMHWKLVQSLFRVLVGLLEIWNKDFY-SRYANVTLENFLSMREQLQOQ
 HPEDVOLIITVDFIGSTVGLAKVSLCSHGAVNWDKSNVIGVASTMAHELGNL
 GMSHDEIIPGCVPEPRGGCINTESIGSKFPIFSRCKIDLESFVTKPOTGCLTK
 VPDNRFVSGVCGNLFVHGEQCDGTPQDCNFCNATTCQLVGAECAGTCCHS
 CKVPAEVCRLSKDDEEFCDFGRKTPEDAFQNGTPCPGGYCFDGSCTLAQQ
 CEDLWPGHVAADSYTISIIPGNGRMVSGSLPSVCRINRCGALYCEGGKPLERS
 FCTFSSNGVCHAGTSGSIIDTELVLQGTKEGKVMGDSQCDLRYVSENCASAK
 NNHGVCNKRCHCKGWAPNCPQRLADVDEQAASTSLPVSVVVVLVLAAMVIV
 AGIYLYKAPROIQRRVAEKPISGLSNPLFYTRDSSLPAKRPDPSETVSTNQPPR
 PIVK"

CDS

QY 141 GluProLeuSerProIleHisArgTyrGlyGlnGluHisAlaLeuTyrLysTyrAsnPro 160
 Db 455 GAGCCCTCTGATCTGATGAGAG---GGCAA---CATGGATGATCAGGCAAGCAT 508
 QY 161 TrpGluLysAsnTyrTrpSerThrTrpGlyMet----- 171
 Db 509 CTGCAACAGAGGC-TGGGACCTGTGGGTCAAAGATACCAACCTGAATGACCTAGGGCC 567
 QY 172 TrpGlyValLeuTrpAlaHisTrpLeuGlnGluAsnIleAlaLeuProAlaThrLysLeu 191
 Db 568 TCGGGCAT-----AGAAATATACAGGGCTCAGGCACCGAACTG 606
 QY 192 -ValLysLeuLysTrpArgLysValGlnGluHisGluLysTyrIleGluTyrTrpLeuVa 211
 Db 607 GCTG-----ATACCCAGAGAAACCCGCTATGTGGAGTGTATGTGGT 648
 QY 211 LLeuTrpAsnGlyGluLysArgTyrAsnGluAsnGlnTrpGluLysArgLysArgVa 231
 Db 649 TGCAGACAGCAAGAGTTCAGAAAGTTGGGAGCAGAGAG---GCCGTGGCCAGCGAGT 705
 QY 231 LTrpGluMetAlaAsnTyrValAsnMetLeuTyrLysLysLeuAsnThrHisAlaLe 251
 Db 706 SCTGGAGGTGTAAACACACGTGCAGCAAGCTTTATCAGAACTCAGTTCCGAGTGTCT 765
 QY 251 uValGlyMetGluIleTrpThrTrpLysTrpLysIleLysIleThrProAsnLysTr 271
 Db 766 GTTGGGCTGAGATCTGGAC---AAGGACAAATTTACATCAGCCGCTATGCCAAGT 822
 QY 271 pThrLeuGluAsnTrpSerLysTyrArgGlySerValLeuSerArgLysArgHisTr 291
 Db 823 GACACTGGAGAACTTCTCTCTCGAGGAAACAGCACTTGCAGGGGAGCAGCCACATGA 882
 QY 291 pIleAlaGlnLeuIleThrAlaThrGluLeuAlaGlyThrThrValGlyLeuAlaTrpMe 311
 Db 883 CAACGTCAACTATCACGGGGGTGATTTCAITGGGAGCACTGTGACATGGGTAGGT 942
 QY 311 tSerThrMetTrpSerProTyrSerValGlyValValGlnTrpHisSerTrpAsnLeu 331
 Db 943 GTCTGCCCTGTGTTCCTGTCACCTCCGAGCTGTGAAATCAGGACCACTCCAAAGAACTCCAT 1002
 QY 331 uArgValAlaGlyThrMetAlaHisGluMetGlyHisAsnTrpGlyMetTrpHisTrpTr 351
 Db 1003 TGGTGTAGCTCCACCATGSCCATGAGTGGGCCACCAACCTGGGCATGACCCAT----- 1057
 QY 351 pTyrSerTrpLysTrpProSerThrIle-----TrpVa 362
 Db 1058 ----GATGAGACATTCAGAGATCTACTCTCTGACACCGAGGGGTGGTGGTGCAT 1113
 QY 362 lMetTrpLysAlaLeuSerTrpTyrIleProThrTrpTrpSerSerTrpSerArgLeuSe 382
 Db 1114 CATGACCGAAAGCATCGCTCCAAAGTTCCTCCAGGATATTCAGCAGGTGTAGCAAGATGA 1173
 QY 382 rTyrTrpLysTrpTrpGluTrpLysLeuSerAsnTrpLeuTrpAsnAlaProLeuProTh 402
 Db 1174 CCTAGAGTCATTCGTGCAAAACCCAGACAGCGCTGCTGACCAATGTTCCAGATGTCAA 1233
 QY 402 rTrpIleIleSerThrProIleTrpGlyAsnGlnLeuValGluMetGlyGluTrpTrpTr 422
 Db 1234 CCGGTTCTGGTGGCCCTCTGTGTGTGMAACCTCTTTTGTGGAGCATGGAGCAGTGTGA 1293
 QY 422 pTrpGlyThrSerGluGluTrpThrAsnIleTrpTrpTrpAlaLysThrTrpLysIleLy 442
 Db 1294 CTGTGGCACACCTCAGGACTGTCAAACCCCTGTCAAATGCCACCACTTCCAGCTGGT 1353
 QY 442 sAlaThrTrpGlnTrpAlaLeuGlyGluTrpTrpGluLysTrpGlnTrpLysLysAlaGl 462
 Db 1354 CAAGGGGCGAGGTGCTCCAGTGTCTGTGTGTCATGAATGCAAGGGAAGGAGCAGCTGG 1413
 QY 462 yMetValTrpArgProAlaLysTrpGluTrpTrpLeuProGluMetTrpAsnGlyLysSe 482
 Db 1414 AGAGGTGTGCTCTCAGTAGAGCAATGTGACCTGGAGGAGTCTCTGTATGCCGGA 1473
 QY 482 rGlyAsnTrpProTrpTrpArgTrpGlnValAsnGlyTrpProTrpHisHisGlyLysGl 502

ORIGIN

Alignment Scores:
 Pred. No.: 6,12e-38 Length: 2277
 Score: 668.00 Matches: 239
 Percent Similarity: 45.63% Conservative: 121
 Best Local Similarity: 30.29% Mismatches: 333
 Query Match: 14.18% Indels: 98
 DB: 11 Gaps: 23

US-10-726-148a-15 (1-775) x AK089169 (1-2277)

QY 7 ProValSerLeu-----LeuLeuSerVal-----AlaValSerAlaIleLys 20
 Db 38 CCATCATGCTTGCTCTGCTCTGCTCAGCGTCTTTA-TGGACACCACTAGCCCTGGACCT 97
 QY 21 GluLeuProGlyValLysTyrGluValValTyrProIleArgGluHisProLeuHis 40
 Db 98 CTTTGGCCCATGAGAAAGATGAAAGTGTGTGGCTCTGGCCCTAGCTGATCCCGC 157
 QY 41 LysArgGluAlaLysGluProGluGlnGlnGluGlnTrpGluThrGluLeuLysTyrLys 60
 Db 158 TCCGGCAGAGCCCTG---CCCTCCACTGGGGCAGTACCACAGAGCTCTGAGCTATGCT 214
 QY 61 MetThrIleAsnGlyLysIleAlaValLeuTyrLysLysLysAsnLysAsnLeuAla 80
 Db 215 CTTGGGACCAAGCGCGCAGTTTTCCCTCCCTGACCTTCGAAAGAACAGGGACCTCTGGGC 274
 QY 81 ProGlyTyrThrGluThrTyrTyrAsnSerThrGlyLysGluIleThrThrSerProGln 100
 Db 275 TCAAGCTACACAGACACTACTAGCTGCCAATGGCTCTGAGGTGACAGCAACTGCAG 334
 QY 101 IleMetTrpTrpTrpTyrTyrGlnGlyHisIleLeuAsnGluLysValSerTrpAlaSer 120
 Db 335 GAGCAGGACCATTCCTCTCAAGAGGCATCTGGAAGGGTACGAGGGGCTCAGCTGCCAGT 394
 QY 121 lIleSerThrTrpArgGlyLeuArgGlyTyrTrpSerGlnGlyTrpGlnArgTyrTrpIle 140
 Db 395 ATTAGCACCTGTGCTGGCCCTCAGGGGCTTTTTCGAGGTGGGTCCACTGCTCCTGATT 454

```

Db      1474 GCCAACATGTCGCCAGATGCTCCAAACAGATGGCACTCCCTGCG-----CCAGGGGG 1527
Qy      502 yHisTrpLeuMetGlyThrTrpProThrLeuGlnGlnTrpThrGluLeuTrpGlyPr 522
Db      1528 CTACTGCTTTCATGGAGCTGTCACCTCCCTGGACAGAGTGGCGGATCTGTGGGGGCC 1587
Qy      522 oGlyThrGluValaTrpLysSerTrpTyrAsnArgAsnGlyGlySerLysTyrGl 542
Db      1588 AGTGTCTGGGTAGCAGCCGACTCTGCTATACCTTTAGCATCCCTCGGGCTGCAATGG 1647
Qy      542 yTyrTrpArgValTrpThrLeuLeuProTrpLysAlaAsnTrpThrMetTrpGl 562
Db      1648 G-----AGGATGACTCTGCAGCTTCCCTCTGTCGAGGATCAACCGGTGG 1698
Qy      562 yLysLeuTrpGlnGlySerTrpAsnLeuProTrpLysGlyArgIleValThrTr 582
Db      1699 AGCGCTGACTGTGAGGGAGC-----CAGAAGCCCTTGAACGCTCTTCGCACTT 1752
Qy      582 pLeuThrTrpLysThrTrpTrpProGluTrpThrSerGlnGluLeuGly----- 598
Db      1753 CTCCTCC-----AACCATGAGTGTGCCATCTCTTGGCACAGGCAGCAA 1797
Qy      599 -----MetValAlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpIl 613
Db      1798 CATGACACCTTTCAGCTGGTATTCGAGGGCACCAGTGGCAGGAGGAAAGGTTGCT 1857
Qy      613 eAsnAlaGluTrpValTrpIleGluLysAlaTyrLysSerThrAsnTrpSerLysTr 633
Db      1858 GGATGGAAGCTGCCAGGACCTC---CGTGTATACAGATCTGAAACCTCTCTGCTAAATG 1914
Qy      633 pLysGlyHisAlaValTrpThrHisGluLeuGlnTrpGlnTrpGluGlyTrpIlePr 653
Db      1915 CAACACCATGGGGTACGCACACCAAGAGAGGAGTGCACCTGTCTCACAGGGCTGGCACC 1974
Qy      653 o-Pro-----TrpTrpTrpTrp-----S 659
Db      1975 ACCCACTGTGTACAGCGGTGGCAGATGATCAGATGACAGCAGCGCTCTACAGCGCT 2034
Qy      659 erSerValValTrpHisTrpSerIleValValGlyValLeuTrpProMetAlaValIle 679
Db      2035 CCCAGTCAGTGTGTGTGGTCTTG-GTGATCCTGGTG-----GCTGCGATGG 2081
Qy      679 rpValValAlaMetValIleArgHisGlnSerSerArgGluLysGlnLysTrpG 699
Db      2082 TCATCGTGGCAGCATCGTATCTACCGAAGGCTCCGAGACAAATCCAGAGGAGGAATG 2141
Qy      699 InArgProLeuSerThrThrGly---ThrArgProHisLysGlnLysArgLysProGlnM 718
Db      2142 TGGCACCCCAAGCCATCTCGGGGCTCTCAACCCCTTATTCTACACAGGGACAGCACC 2201
Qy      718 etValLysA-avalGlnProGlnGluMetSerGlnMetLysProHisValTyrTrpLeup 738
Db      2202 TGCCAGCTAAGACAGGCTCCAGACCTTCTGAG-----A 2237
Qy      738 roValGluGlyAsnGluProPro 745
Db      2238 CAGTTTCTACCAACAGCCCCCA 2260

RESULT 5
BE538260 744 bp mRNA linear EST 09-AUG-2000
LOCUS 6c1062138f1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3448832 5',
DEFINITION mRNA sequence.
ACCESSION BE538260
VERSION BE538260.1 GI:9766905
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 744)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
```

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: Incyte Genomics, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LHAM8424 row: k column: 09
 High quality sequence stop: 614.
 Location/Qualifiers
 source
 1. 744
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3448832"
 /cell_line="MGC36"
 /lab_host="DH10B"
 /note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 1.5 kb. Library prepared by Life
 Technologies."

ORIGIN

Alignment Scores:
 Pred. No.: 1,97e-38 Length: 744
 Score: 665.50 Matches: 154
 Percent Similarity: 83.76% Conservative: 11
 Best local Similarity: 78.37% Mismatches: 30
 Query Match: 14.12% Indels: 3
 DB: 10 Gaps: 1

US-10-726-148A-15 (1-775) x BE538260 (1-744)

Qy 238 VallasnMetLeuTyrLysLysLeuAsnThrHisValAlaLeuValGlyMetGluIleTrp 257
 Db 45 GTTTTGTGACCTTTATAAAGCTCAATCACTCATGTGGCTTAGTGGTAAGAAATCTGG 104
 Qy 258 ThrTrpLysTrpLysIleLysIleThrProAsnAlaSerTrpThrLeuGluAsnTrpSer 277
 Db 105 ACTGACAGAGTAAGATAAGATACCCCAATGACAGCTTCACCTTGGAGATTTTCT 164
 Qy 278 LysTrpArgGlySerValLeuSerArgArgLysArgHisTrpIleAlaGlnLeuIleThr 297
 Db 165 AAATGGAGGGGAGTGTCTCTCAAGAAAGAAAGCGTCATGATATTGCTCAGTTAATCACA 224
 Qy 298 AlaThrGluLeuAlaGlyThrThrValGlyLeuAlaTrpMetSerThrMetTrpSerPro 317
 Db 225 GCAACAGAACTTGTCTGGAAACGACTGTGGGTCT-GCATTTATGTCTACAAATGTTCTCT 283
 Qy 318 TyrSerValGlyValValGlnTrpHisSerTrpAsnLeuLeuArgValAlaGlyThrMet 337
 Db 284 TATTCTGTGGGTGTTCAGGACCAAGCGATATATCTTTAGAGTTGCAGGCAATG 343
 Qy 338 AlaHisGluMetGlyHisAsnTrpGlyMetTrpHisTrpTrpTyrSerTrpLysTrpPro 357
 Db 344 GCACATGAATGGGGCCCAACTTTGGAATGTTTATGACGACTATTCTTTCGAAGTGTCT 403
 Qy 358 SerThrIleTrpValMetTrpLysAlaLeuSerTrpTyrIleProThrTrpTrpSerSer 377
 Db 404 TCTACAATATGTGTGATGCAAAAGCACTGAGCTTCTATATACCCACAGACTTCAGTTCC 463
 Qy 378 TrpSerArgLeuSerTyrTrpLysTrpTrpGluTrpLysLeuSerAsnTrpLeuTrpAsn 397
 Db 464 TGCAGCGCTCTCAGCTATGACAGATTTTGAAGATAAATATCAATATGCTCTCTTAAT 523
 Qy 398 AlaProLeuProThrTrpIleSerThrProIleTrpGlyAsnGlnLeu-ValGluMe 417
 Db 524 GCTCCATGTGTACAGATATCATATCCACTCCAATTTGTGGGAACCAAGTTGGTAGAAAA 583

Db	528	AGCTGGAGAACTTGGATGAACAGGCGACATCTTCAAGGACGACTCCATATGAAGAC	587
QY	164	AsnTyrTrpSerThrTrpGlyMetTrpGlyValLeuTrpAlaHisTrpLeuGlnGlnAsn	183
Db	588	CAGGAGNAGCAACTGTGTGGATGATGGCTTCACAGGCTGCATCAGAC	647
QY	184	IleAlaLeuProLathrLysLeuVal	192
Db	648	GTGGCCCCCTGCACAGGTGATT	674
RESULT 8	BQ045321/c		
LOCUS	DEFINITION	697 bp mRNA linear EST 21-FEB-2003	
	ACCESSION	UI-CF-EN1-aei-b-05-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone	
	VERSION	UI-CF-EN1-aei-b-05-0-UI 3', mRNA sequence.	
	KEYWORDS	BQ045321.1 GI:19796516	
	SOURCE	EST.	
	ORGANISM	Homo sapiens (human)	
REFERENCE	AUTHORS	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
	TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
		1 (bases 1 to 697);	
		Bonaldo,M.F., Lennon,G. and Soares,X.B.	
		Normalization and subtraction: two approaches to facilitate gene	
		discovery	
		Genome Res. 6 {9}, 791-806 (1996)	
	MEDLINE	97044477	
	PUBMED	8889548	
COMMENT	CONTACT	McCrack, PB	
		University of Iowa	
		2024 University of Iowa Med Labs, Iowa City, IA 52242, USA	
		Tel: 319 356 4866	
		Fax: 319 356 7171	
		Email: paul-mccray@uiowa.edu	
		Tissue Procurement: Dr. M. J. Welsh, University of Iowa	
		cDNA Library preparation: Dr. M. Bento Soares, University of Iowa	
		cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa	
		DNA Sequencing by: Dr. M. Bento Soares, University of Iowa	
		Clone Distribution: Researchers may obtain clones from Research	
		Genetics (www.resgen.com) or from Oper Biosystems	
		(www.openbiosystems.com).	
		The following repetitive elements were found in this cDNA	
		sequence: 1-45, >POLY A#Simple_repeat (matched complement)	
		Seg primer: M13 FORWARD	
		polyA-vec	

TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS					
6hr to LPS 24h					
TAG LIB=UI-CF-ENI					
TAG_SEQ=CTGCTCAGGT"					
ORIGIN					
Alignment Scores:					
Pred. No.:	4,97e-32	Length:	697		
Score:	577.00	Matches:	125		
Percent Similarity:	83.23%	Conservative:	9		
Best Local Similarity:	77.64%	Mismatches:	23		
Query Match:	12.25%	Indels:	4		
DB:	12	Gaps:	1		
US-10-726-148A-15 (1-775) x BQ045321 (1-637)					
Qy	576	LysGlyArgILEValTrpTrpLeuThrTrpLysThrTrpTrpProGluTrpThrSerGln	595		
Db	696	AAAGGCGATAGTGACTTTTCTGCACATGTAAACATNTGATCCTGAAGACACAAGTCAA	637		
Qy	596	GluILEGlyMetValAlaAsnGLYThrLysTrpGlyTrpAsnLysValTrpIleAsnAla	615		
Db	636	GAATAAGCCATGGTGGCCAATGGCAACTAAGTGTGGCGCATACAAGGTTTGCATTATATGCA	577		
Qy	616	GluTrpValTrpIleGluLysAlaValLysSerThrAsnTrpSerSerLysTrpLysGly	635		
Db	576	GAATGTGTGATATTGAGAAAGCTACAAATCAACCAATTCTTATCTTAAGTGCAAGA	517		
Qy	636	HisAlaValTrpTrpHisGluLeuGlcTrpGlnTrpGluGluGlyTrpIleProProTrp	655		
Db	516	CATGCTGTGTGTGACCATGAGCTCCAGTGTCAATGTGAGGAAGGATGGATCCCTCCCAC	457		
Qy	656	TrpTrpTrpSerValValTrpHisTrpSerIleValValGlyValLeuTrpProMet	675		
Db	456	TGGATGACTTCCTCAGTGTCTTCCACTTCTTCCCTGTGGTGGGTGCTGTTCCTCAATG	397		
Qy	676	AlaValIleTrpValValValAlaMetValIleArgHisGlnSerSerArgGluLysGln	695		
Db	396	CGCGTCATTATTTTGTGTGTGTGTGTATGGTATATCCGCCACCAGAGCTCCAGAGAAAGCAG	337		
Qy	696	LysLysTrpGlnArgProLeuSerThrThrGlyThrArgProHisLysGlnLysArgLys	715		
Db	336	AAGAAAGATCAGAGGCCACTATTATCACCTGGCCACCGGCCACACAAACAGAGAGGAAA	277		
Qy	716	ProGlnMetValLysAlaValGlnProGlnGluMetSerGlnMetLysProHisValTyr	735		
Db	276	CCCCAGATGGTAAAGGCTGTTCACCCCAAGAGGTGAAC-----TATATACTC	229		
Qy	736	TTP 736			
Db	228	TGG 226			
RESULT 9					
BB629805					
LOCUS					
DEFINITION					
musculus cDNA clone 9930012G10 5', mRNA sequence.					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
Mus musculus					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;					
1 (bases 1 to 667)					
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,					
Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,					
Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,X.,					
Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,					
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,					
Sogabe,Y., Suzuki,H., Tagami,M., Tegawa,A., Takahashi,F.,					
Takeda,Y., Tanaka,T., Toyota,T., Muramatsu,M. and Hayashizaki,Y.					
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)					
TITLE					

1454 AGAGGACTGATTACGGTCAGCAGCAACCTCAGCTACGTATCATCGAGCCCTCCTCGACAGC 513
144 -----SerProileHisArgTrpGly 150
514 AAGGCCCAACACCTATTATTACAGATCTGAACATCTCAAGCCGCCCGCGGAAACTGTGGG 573
151 GlnGluHisAlaLeuTrpLysTrpAsnProTrpGluLysAsnTrpTrpSerThrTrpGly 170
574 TTCGAGCACTCCAGCCACCCAGCCAGGAGCTGG----- 606
171 MetTrpGlyValLeuTrpAlaHisTrpLeuGlnGlnAsnLeuAlaLeuProAlaThrLys 190
607 -----GCTCTTCAGTTTACACA 524
191 LeuValLysLeuLysTrpArgLysValGlnGluHisGlu-----LysTyrIle 206
625 CAGACCAAGAACCGACCTCGCAGGATGAAGAGGAGATTTAAACTCCATGAAGTATGTG 684
207 GluTyrTrpLeuValLeuTrpAsnGlyGluTrpLysArgTyrAsnGluAsnGlnTrpGlu 226
685 GAGCTTTACCTCGTGGCTGATTATTAGAGTTTCAGAGAAATCGAGAGACCCAGGACGCC 744
227 IleArgLysArgValTrpGluMetAlaAsnTyrValAsnMetLeuTyrLysLysLeuAsn 246
745 ACCAAACACAAGCTCATAGAGATGCCCACTATGTTGATAGTTTACCGATCTTGAAC 804
247 ThrHisValAlaLeuValGlyMetGluLeuTrpTrpLysTrpLysLysIleThr 266
805 ATCCGGATGTCTCTGTGGCTTGGAGTGTGGACCCAGCGGAAATGTGTGAATTTCA 864
267 ProAsnAlaSerTrpThrLeuAsnTrpSerLysTrpArgGlySerValLeuSerArg 286
865 GAGATCCATATTCTACCTCTGGTCCCTTCTCAGTTGGAGCGCAAGCTGCTGCC--- 921
287 ArgLysArgHisTrpIleAlaGlnLeuIleThrAlaThrGluLeuAlaGlyThrTrpVal 306
922 CAGAAGTACCATGACAAACGCCCAATTAATCAGGCGATGCTCTCCAGCGCACCAACATC 981
307 GlyLeuAlaTrpMetSerThrMetTrpSerProTyr-----SerValGlyValGlnTrp 325
982 GGCTGGCCGCCCTCATGGCCATGCTGTGTACCATGCTGTGGAGAGTCAACATGGAC 1041
326 HisSerTrpAsnLeuArgValAlaGlyThrMetAlaHisGluMetGlyHisAsnTrp 345
1042 CACTCCGGAGATGCATTGGGTGGCTGCCACCATGGCCCGAGATGGCCCAACATTT 1101
346 GlyMetTrpHisTrpTrpTrpTrpTrpProSerThrIleTrpValMetTrpLys 365
1102 GGCATGACCCAT-----GATTCTGCAGATTGCTCGGCCAGT 1140
366 AlaLeuSerTrpTyrIleProThrTrpTrpSerSerTrpSerArgLeuSer----- 382
1141 GCGGCTGA-TGGTGGGTGCATCATGGCAGCTGCCACTGGGCACTCTTCCCAAGTGT 1199
383 TyrTrp-----LysTrpTrpGluTrpLys 390
1200 CAATGGATGCAACAGAGGGAGCTGGACAGGTATCTGCAGTCAGTGGTGGT----- 1250
391 LeuSerAsn-TrpLeuTrpAsnAlaProLeuProThrTrpIleIleSerThrProIleTr 410
1251 -----AATGTGTCTCTCCACATGCCAGACACAGGATGTTGATGGAGCGCGAGGTG 1304
410 pGlyAsnGlnLeuValGluMetGlyGluTrpTrpTrpTrpGlyThrSerGluGluTrpTh 430
1305 TGGGAACCGGTATCTCGGAAGATGGGGAAGAGTGTGACTGTGGAGAAGAGAGATGTAA 1364
430 rAsnIleTrpTrpTrpAlaLysThrTrpLysIleLysAlaThrTrpGlnTrpAlaLeuG 450
1365 CACCCCTGCTGCATGCTCTAATTGATACCTGAGGCCGGGGCGGAGTGTGCTCACGG 1424
450 yGluTrpTrpGluLysTrpGlnTrpLysLysAlaGlyMetValTrpArgProAlaLysTr 470

1425 CTCCTGCTCCACCACTGTAAGCTGTGTGCTCTCTGGGACCTGTGTGGGAGAGAGCCAG 1484
470 pGluTrpTrpLeuProGluMetTrpAsnGlyLysSerGlyAsnTrpProTrpTrpArgTr 490
1485 GCAGTGTGACCTCCCGGAGTTCTGTACGGCGCAAGTCTCCGCCACTGCCCTACCAACTTCTA 1544
490 pGlnValAsnGlyTrpTrpTrpHisHisGlyLysGlyHisTrpLeuMetGlyThrTrpPr 510
1545 CCAGATGATGTATACCCCTGTGAGGCGCGCCAGGCTACTGCTCAACAGCGCATGTGCCT 1604
510 oThrLeuGlnGlnTrpThrGluLeuTrpGlyProGlyThrGluValAlaTrpLysSe 530
1605 CACTTACCAGAGCAGTCCAGAGCTGTGGGACCCGAGCCGACCTGCCCTGACCT 1664
530 rTrpTrpAsnArgAsnGluGlyGlySerLysTyrGlyTyrTrpArgArgValTrpTrpTh 550
1665 CTGCTTCAGAGAGGT-GAATGTGGC-----AGGAGACACCTTTGGAA 1705
550 rLeuIleProTrpLysAlaAsnTrpThrMetTrpGlyLysLeuTrp----- 565
1706 ACTG-----TGGAAA-----GGTCATGAATGTGAACACAGGAAGT 1741
566 -----Trp----- 566
1742 GCAACATGAGAGATGCGAAGTGTGGGAAGATCCAGTGTGAGCTCTGAGGCCCGGCC 1801
567 -----GlnGlyGlySer-----TrpAsnLeuProTrpLysGlyArgIleVa 580
1802 TGGAGTCCAAACGCGGTGCCATTGACCACTATTCATCATGAATGGGA-GGCAGATCCAG 1860
580 lThrTrpLeuThrTrpLysThrTrpTrpProGluTrpThrSerGlnGluIle-----G 598
1861 TGCCGGGCGACCCAGCTTACCAG-GGCTCTGAGGAGGGGTGACATCTGNACCCAGG 1919
598 yMetValAlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpIleAsnAlaGluTrpVa 618
1920 GCTGTGTATGACTGGAACCAAGTGTGGCTACCAACATATTGCTTTGAGGGCGAG--TG 1976
618 lTrpIleGluLysAlaTyrLysSerThrAsnTrpSerSerLysTrpLysGlyHisAlaVa 638
1977 CAGGAACACCTCTCTTTTGAACCTGAAGGCTGTGGGAAGAACTGCAATGGCCATGGGT 2036
638 lTrpTrpHisGluLeuGlnTrpGlnTrpGluGluGlyTrpIleProTrpTrpTrpTr 658
2037 CTGTAAACAAACACAGAACTGCCACTGCTCGCGGCTGGGCCCGCCGCC-----TT 2087
658 pSerSerValValTrpHisTrpSer----- 666
2088 CTGCAACACACCGGCGCCACGGGGCGAGTATCGACAGTGGGCGCTATGCCCTCGAGAGTGT 2147
667 -----IleValValGlyValLeuTrpProMetAlaValIleTrpValValAlaMe 684
2148 GGTCTCTGTGTGTAGTGGAGTGTG-----GTGGCCATCTTGTGGTGTGGCGTCTCT 2198
684 tValIleArgHisGlnSerSerArgGluLysGlnLysLysTrpGlnArgProLeuSerTh 704
2199 CATGCTGATGTACTACTCTGTCGACAGACAGAACAACTAGCCCACTCAAGCCCTCAGC 2258
704 rThrGlyThrArgProHisLysGln----- 712
2259 TCTCCCTTCCAAGCTGAGGCAACAGTTCAGTTGTCTCTTCCAGGGTTCCTCAGAACAGCGG 2318
713 -----LysArgLys----- 715
2319 SACTGGTCAATGCCAACCCAACTTTCAAGCTGCAGACGCCGCCAGGCAAGAGGTGAT 2378
716 -----ProGlnMetValLysAlaValGlnProGlnGluMetSerGlnMetLysProHi 733
2379 CAACACTCCGGAATCTCGCG-----AAGCCCTCCAGGCTCTCTCCCGGCCCTCC 2432
733 sValTrpTrpLeuProValGluGlyAsnGluProProAla 746
2433 AGATTAT-----CTGCGTGGTGGTCTCCCACTGCA 2463

```

RESULT 12
BB660137
LOCUS
DEFINITION
BB660137 RIKEN full-length enriched, 13 days embryo lung Mus
MUSCULUS cDNA clone D43003C21 5', mRNA sequence.
ACCESSION
BB660137
VERSION
BB660137.1 GI:16493958
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 634)
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T.,
Hara, A., Hiramoto, K., Ito, F., Ishii, Y., Ito, M., Kawai, J.,
Konno, H., Konda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,
Ono, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
Takeda, Y., Tanaka, T., Toyama, T., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watabiki, M., Yoceda, Y., Ishikawa, T., Orawa, K., Tanaka, T.,
Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.
and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,
Sugahara, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I.,
Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
Hayashizaki, Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
FEATURES
location/Qualifiers
1..634
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="D43003C21"
/tissue_type="lung"
/dev_stage="13 days embryo"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 13 days embryo
lung"
/note="Site 1: Sali; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken

```

```

contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5',
GAGAGAGAGCGCGCCCACTCGAGTCTTTTCTTTTCTTTT 3'], cDNA was
prepared by using thermostable thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5',
GAGAGAGAGATCTCGAGTTAATTAATTAATCCCCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I.*

ORIGIN
Alignment Scores:
Pred. No.: 2,15e-29 Length: 634
Score: 540.00 Matches: 120
Percent Similarity: 73.96% Conservative: 22
Best Local Similarity: 62.50% Mismatches: 39
Query Match: 11.46% Indels: 11
DB: 4
Gaps: 4

US-10-726-148A-15 (1-775) x BB660137 (1-634)

QY 5 LeuLeuProValSerLeuLeuLeuSer---ValAlaValSerAlaIleLysGluLeuPro 23
DB 56 CTTCTGCTAGTCTCTTCTCTTCTTCTCCAGTTCAGTAAGTGCATTAAGAAGACTCCCT 115
QY 24 GlyValLysLysTyrGluValValTyrProLeuArgLeuHisProLeuHisLysArgGlu 43
DB 116 AAAGCCAGAAATATGAAGTGCTTTATCCCATAGACTTCATCCATTCGCTAAAGAGAG 175
QY 44 AlalysGluProGluGlnGlnGluTrpGluTrpGluLeuLysTyrLysMetThrIle 63
DB 176 ACCCAAGAGCCAGAGCCAAAGGAAACATTGAACTGAGCTAAGGTACAAATGACAGTA 235
QY 64 AsnGlyLysIleAlaValLeuTyrLysLysLysAsnLysAsnLeuLeuAlaProGlyTyr 83
DB 236 AATGGAAGGTGCTGCTGCTGCTATCTGAGAGAGAACACCAAGCTCCTTCGCGCTGACTAC 295
QY 84 ThrGluThrTyrTyrAsnSerThrGlyLysGluLeuThrThrSerProGlnIleMetTrp 103
DB 296 TCGGAAACATACTATAATTCAGTGGAAACCAAGGTCCACCAAGCCCGCAATCATGGAT 355
QY 104 TrpTrpTyrTyrGlnGlyHisIleLeuAsnGluLysValSerTrpAlaSerIleSerThr 123
DB 356 AGCTGTACTACCAAGGACACATCGTAATAGAGAAAGTTCTCGAGCAGCATCAGACCC 415
QY 124 TrpArgGlyLeuArgGlyTyrTrpSerGlnGlyTrpGlnArgTyrTrpIleGluProLeu 143
DB 416 TGTCAGGAGACTACGCGGTTCATCATCAGTCAAGAGATGAAAGTATTTTATCGAACCTTTG 475
QY 144 SerProIleHisArgTrpGlyGlnGluHisAlaLeuTrpLys-TyrAsn----- 159
DB 476 AGCTCGGAGAACTTGGATGAAACAGCAGCATGCACTCTTCAAGGAGCAGACTCCAATGAAGA 535
QY 160 -ProTrpGluLysAsnTyrTrpSerThrTrpGlyMetTrpGlyValLeuTrpAlaHisTr 179
DB 536 ACCAGAGAGAGCAGCACTGTGGT-----GGAATGATGCCGTATATGGCTTCAAGGGG 589
QY 179 pLeuGlnGlnAsn-----IleAlaLeuPro 187
DB 590 CTTGCATCAAGGAGCTGGGCGNCCTTCCTTGCCTGCA 623

RESULT 13
BC026542
LOCUS
DEFINITION
BC026542 Mus musculus, decysin, clone IMAGE:4990623, mRNA.
ACCESSION
BC026542.1 GI:20072334
KEYWORDS
HTC.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

```

REFERENCE 1 (bases 1 to 2453)
 AUTHORS Strausberg,R.
 TITLE Direct Submission
 JOURNAL Submitted (02-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 REMARK NIH-XGC Project URL: http://mgc.nci.nih.gov
 COMMENT Contact: MGC help desk
 Email: cgspbs-remail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Prepared by: The I.M.A.G.E. Consortium (LInL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: http://www-shgc.stanford.edu
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. X.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LInL at: http://image.llnl.gov
 Series: IRAC Plate: 58 Row: h Column: 15
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 10946873
 This clone has the following problem: frame shifted.

FEATURES
 source
 1..2453
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4990623"
 /tissue_type="Colon, normal. 5 month old male mouse."
 /clone_lib="NCI CGAP_C024"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

ORIGIN
 Alignment Scores:
 Pred. No.: 2,65e-27 Length: 2453
 Score: 522.50 Matches: 148
 Percent Similarity: 52.21% Conservative: 88
 Best Local Similarity: 32.74% Mismatches: 178
 Query Match: 11.09% Indels: 38
 DS: 11 Gaps: 8

US-10-726-148A-15 (1-775) x BC026542 (1-2453)

QY 10 LeuLeuSerValAlaValSerAlaIleLysGluLeuProGlyValLysLysTyrGlu 29
 Db 119 CTTATCATCAATCCAGTGTATAGTACACAGTTAAACCCCGAGTTAAAGCCTCACGAA 178
 QY 30 ValValTyrProIleArgLeuHisProLeuHisLysArgGluAlaLysGluProGluGln 49
 Db 179 ATAGTTCGTCTCAAAAAACTA---CCCATTTACAGAAAAGAGG33CTTGAGAACACACAG 235
 QY 50 -----GlnGluGlnTyrGluThrGluLeuLysTyrLysMetThrIle 63
 Db 236 ACAGAGAGATATGCAAGAGGAAAAATATCCACAGAAAGTTCAATATCATATCATATACTA 295
 QY 64 AsnGlyLysIleAlaValLeuTyrIleLysLysAsnLysAsnLeuLeuAlaProGlyTyr 83
 Db 296 AATGGAGAGAAATGTTTTTTCATCAAGAGAACCAAGCATCTTCGGGCCACACTAC 355
 QY 84 ThrGluThrTyrTyrAsnSerThrGlyLysGluIleThrThrSerProGlnIleMetTyr 103
 Db 356 ACTGAACAATCTTACTACCCAGGAGAGAGAGACACACACTCTCAGGATGTGAA 415
 QY 104 TrpTrpTyrGlnGlyHisIleLeuAsnGluLysValSerTyrAlaSerIleSerThr 123
 Db 416 CCCTGCTACTATGAAGGCCACATCAAAATCAAGGG33CTCTCGCCAGGATCAGTACC 475

QY 124 TrpArgGlyLeuArgGlyTyrTrpSerGlnGlyTyrGlnArgTyrTrpIleGluProLeu 143
 Db 476 TGTGATGGTTCAGAGGGTACTTCACATCGTGAATCAGAGATATCAATAAAGCCCTCTA 535
 QY 144 SerProIleHisArgTyrGlyGlnGluHisAlaLeuTrpLysTyrAsnProTyr----- 161
 Db 536 CAAAGCACAGATGAG-----GGAGAGCATGCTGTTCTCCCATACAGC---TGAAGGGA 586
 QY 162 -----GluLysAsnTyrTrpSerThrTrpGlyMetTyr 172
 Db 587 CAAGACACAGTTTCATGCAAGGATGCTGAGAACAGTT----- 625
 QY 173 GlyValLeuTyrAlaHisTrpLeuGlnAsnIleAlaLeuProAlaThrLysLeuVal 192
 Db 626 -----GTCAGGAACGAGGCCATCTTCGAACTTCAGAGGTCACTC 664
 QY 193 LysLeuLysTyrArgLysValGlnGluHisGluLysTyrIleGluTyrTrpLysValLeu 212
 Db 665 AAAAACCCCAATGAGACTTACTTCAAGGACAGAGTACATTGCTCTCTTTGGTGTCTG 724
 QY 213 TrpAsnGlyGluTyrLysArgTyrAsnGluAsnGlnTrpGluIleArgLysArgValTyr 232
 Db 725 GATAATGCTATTTAATAGCTGTATATG3GAAATGTGACTCAGATGAGAACCTTTCTGTTC 784
 QY 233 GluMetAlaAsnTyrValAsnMetLeuTyrLysLysLeuAsnThrHisValAlaLeuVal 252
 Db 785 AAGTGTGAAATCTGCTCAATATGATTTATAAACCAATAATATTCAAGTGTCTTTGGTG 844
 QY 253 GlyMetGluIleTyrThrTrpLysTyrLysIleLysIleThrProAsnAlaSerTyrThr 272
 Db 845 GGCATGGAGATCTGCTCTGACCCAGGACAGATAAAGGTGGAGCCCAACCTTGGTGCACG 904
 QY 273 LeuGluAsnTyrSerLysTyrArgGlySerValLeuSerArgArgLysArgHisTyrIle 292
 Db 905 TTCACACACTTCATGAGATGGCATTTATTCTAACCTGGGGAAAAAGG---ATCCATAACAC 961
 QY 293 AlaGlnLeuIleThrAlaThrGluLeuAlaGlyThrThrValGlyLeuAlaTyrMetSer 312
 Db 962 GCCCAGCTTCTCAGTGGAGTAGCTTCGCCCATGGAGCTGGGAATGGCAGCTGGGAAT 1021
 QY 313 ThrMetTyrSerProTyrSerValGlyValValGln-TrpHisSerTyrAsnLeuLeuAr 332
 Db 1022 TCCTTTTCCTACTACATCGTCTGTCTGTAATGAGGCTAAAAAAGAAATAATGTGGC 1081
 QY 332 gValAlaGlyThrMetAlaHisGluMetGlyHisAsnTrpGlyMetTyrHisTyrTyr 352
 Db 1082 TCTTGTGCACTGATGTACATGATGTTGGCCACGCCCTTGGTATGAAGATGTTCCATA 1141
 QY 352 rSerTyrLysTyrProSerThrIleTyrValMetTyrLysAlaLeuSerTyrTyrIlePr 372
 Db 1142 TTATACCAAGTGTCCCTCTGCGAGCTGTGTGATGATCAGTACCTGAGTTCAAAATTCCT 1201
 QY 372 cThrTyrTrpSerTyrTrpSerArgLeuSerTyrTrpLysTyrTrpGluTyrLysLeuSe 392
 Db 1282 CAAGATTTAGTACCGTCAGCCGTTCTCAATTTCAAGGGTTTCTCTCATCTAGAAATGC 1261
 QY 392 rAsnTrpLeuTrpAsnAlaProLeuProThrTyrIleIleSerThrProIleTyrGlyAs 412
 Db 1262 AAGGTGCTGCTGCTGGCACCTGACCCAAAAACATAATAAAA---CCAACATGGGGAA 1318
 QY 412 nGlnLeuValGluMetGlyGluTyrTrpTyrTrpGlyThrSerGluGluTyrThrAsnI 432
 Db 1319 CCAAGTCTCGATGTGGAGAGAGTGTGACTGTGGTCTCTCTCTGAGGAATGTACCAATCT 1378
 QY 432 eTyrTrpTrpAlaLysThrTrpLysIleLysAla 443
 Db 1379 TTGTGTGAGGCCCTGCATCATGTAGCTGAAATCT 1412

RESULT 14
 BY721112 675 bp mRNA linear EST 17-DEC-2002
 LOCUS BY721112 RIKEN full-length enriched, adult male epididymis Mus
 DEFINITION musculus cDNA clone 9230028017 5', mRNA sequence.


```

Qy 453 TtpGluLysTrpGlnTrpLysAlaGlyXetValTtpAisqProAlaLysTrpGlnTrp 472
Db 422 TGTGAGAAATCCCACTTAAAAACCTGGGTGTGTCCAGAGCAGCAAGATGAGT 481
Qy 473 TtpLeuProGluMetTtpAsnGlyLysSerGlyAsnTrpProTtpTrpArgTrpGlnVal 492
Db 482 GATCTGCTGAAGTGTGTGATGTAATCCAGCCACTGCCAGGGTGACAGATTCAGATC 541
Qy 493 AsnGlyTrpProTtpHisGlyLysGlyHisTrpLeuMetGlyThrTrpProThrIeu 512
Db 542 AATGGCTCCCTTCGCCAAATGGCGATGTTACTGTTTGAAGGGCAATGTCCACCCCTG 601
Qy 513 GlnGluGlnTrpThrGluLeuTrpGlyProGlyThrGluValAlaTtp 528
Db 602 CAGCAGCAGTGCATGCACATGTGGGTCCAGAACACAGGGTGCATAC 649

RESULT 15
BQ028158/c
LOCUS
DEFINITION
  UI-H-CO0-ari-h-03-0-UI.s1 NCI CGAP_Sub9 Homo sapiens CDNA clone
  IMAGE:3106779 3', mRNA sequence.
ACCESSION
  BQ028158
VERSION
  BQ028158.1 GI:19763437
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1. (bases 1 to 357)
REFERENCE
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  Contact: Robert Strausberg, Ph.D.
  Email: cgaps-remail@nih.gov
  Tissue Procurement: Dr. Jose Mercuende
  cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: Clone distribution information can be found
  through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
  Seq primer: M13 FORWARD
  POLYA=Yes.
FEATURES
  Location/Qualifiers
    1..357
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:3106779"
      /tissue_type="mixed"
      /dev_stage="mixed"
      /lab_host="DH10B (Life Technologies)"
      /clone_lib="NCI CGAP Sub9"
      /notes="vector: pT7T3-Pac (Pharmacia) with a modified
      polylinker; Site 1: EcoR I; Site 2: Not I; tissues:
      Cholonic mucosa with Crohns disease, Cholonic mucosa with
      ulcerative colitis, Fetal thymus, Cervix, Cervical
      adenosquamous carcinoma, Ligament cells, Prostate
      carcinoma, Bladder carcinoma, Brain oligodendrocyte;
      NCI CGAP Sub9 is a subtracted cDNA library constructed
      according to Bonaldo, Lemmon and Soares, Genome Research,
      6:791-806, 1996. First strand cDNA synthesis was primed
      with an oligo-dT primer containing a Not I site. Double
      stranded cDNA was ligated to an EcoR I adaptor, digested
      with Not I, and cloned directionally into pT7T3-Pac
      vector. The oligonucleotide used to prime the synthesis of
      first-strand cDNA contains a library tag sequence that is
      located between the Not I site and the (dT)18 tail. The
      sequence tags for this library are CGTC, AAGC, GGCC,
      GGAG, TAGC, ATGG, AGACA, ATCAC. For additional
      information, contact: Bento Soares, bento-soares@uiowa.edu
      TAG_TISSUE=Cervix
      TAG_LIB=UI-H-CO0

```

TAG_SEQ=GGGCC"

ORIGIN

Alignment Scores:

Pred. No.: 1.54e-26 Length: 357
 Score: 496.00 Matches: 103
 Percent Similarity: 93.81% Conservatives: 3
 Best local Similarity: 91.15% Mismatches: 7
 Query Match: 10.53% Indels: 1
 DB: 12 Gaps: 0

US-10-726-148A-15 (1-775) x BQ028158 (1-357)

Qy 25 VallyLysTyrGluValValTyrProLeuArgLeuHisProLeuHisLysArgGluAla 44
 Db 350 GTGAAGAAAGTATGAAGTGGTTTATCTATAGACTTCATCCACTGCATAAAAGAGAGGCC 291
 Qy 45 LysGluProGluGlnGlnGluGlnTrpGluThrGluLeuLysTyrLysMetThrIleAsn 64
 Db 290 AAGAGCCAGAGCAACAGAACATTTGAACTGAATTAAGTATAAAATGACAATTAAT 231
 Qy 65 GlyLysIleAlaValLeuTyrLysLysAsnLysAsnLeuLeuAlaProGlyTyrThr 84
 Db 230 GGAATAATTCAGTCTTTTATTGAA-AAAAACAGAACCTCTTGCACCGCTACACG 172
 Qy 85 GluThrTyrAsnSerThrGlyLysGluIleThrSerProGlnIleMetTrpTrp 104
 Db 171 GAAACNTATTATAATTCACCTGGAAGAGGAGATCCACCAAGGCCCAAAATTATGGATGAT 112
 Qy 105 TrpTyrTrpGlnGlyHisIleLeuAsnGluLysValSerTtpAlaSerIleSerThrTrp 124
 Db 111 TGTATTATCAAGGACATATTCTTAAATAAAGGTTTCTGACGTAGCATCAGCATGT 52
 Qy 125 ArgGlyLeuArgGlyTyrTyrTrpSerGlnGlyTrpGlyArg 137
 Db 51 AGGGGTCTAAGGGGCTACTTCAGTCAGGGGGGATCAAAAA 13

Search completed: June 10, 2004, 00:58:08

Job time : 4973 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein: - nucleic search, using frame_plus_p2n model
Run on: June 3, 2004, 20:20:30 ; Search time 749 Seconds
(without alignments)
4395.666 Million cell updates/sec

Title: US-10-726-148A-15
Perfect score: 4712
Sequence: 1 MLQGLLPSVLLLSVAVSAIK.....PPTVKNPMSPTKWSNPXA 775

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Xgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ps2r:model -DEV=xlh
-Q=/cgc2_1/USP70_spool/US10726148/runat_09062004_161328_5270/app_query.fasta_1.967
-DB=N_Geneseq_25Jan04 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORW=ext -HEAPSIZ=500 -MINLEN=C -MAXLEN=2000000000
-USER=US10726148 -CGC 1_1_513 -runat_09062004_161328_5270 -NCPU=6 -ICPU=3
-NO MWAF -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_25Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001as:*
5: Geneseq2001bs:*
6: Geneseq2002as:*
7: Geneseq2003as:*
8: Geneseq2003bs:*
9: Geneseq2003cs:*
10: Geneseq2004as:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	3128	66.4	2325	3 AAZ89683	Aaz89683 Human ADA
2	3128	66.4	2828	3 AAZ89726	Aaz89726 Human ADA
3	3128	66.4	3179	4 ABA09173	Abao9173 Human eMD
4	2132	45.2	1620	3 AAZ89673	Aaz89673 Human ADA
5	2129	45.2	2560	3 AAZ89725	Aaz89725 Human ADA
6	2098	44.5	1590	2 AAT63124	Aat63124 Human met
7	2098	44.5	2057	2 AAT63123	Aat63123 Human met
8	1610	34.2	2047	4 AAZ89683	Aaz89683 Human ADA

9	1610	34.2	2047	4 AAZ89683	Aaz89683 Human ADA
10	1510	34.2	2047	4 AAZ89683	Aaz89683 Human ADA
11	1576	33.4	2049	7 ABX90528	Abx90528 Human cDN
12	854	18.1	1652	3 AAZ89683	Aaz89683 Human ADA
13	820	17.4	1704	4 AAZ89683	Aaz89683 Human ADA
14	820	17.4	1704	4 AAZ89683	Aaz89683 Human ADA
15	763	16.2	2334	6 ABK10644	Abk10644 Pigmy rat
16	759.5	16.1	2635	4 AAZ89683	Aaz89683 Human ADA
17	750.5	15.9	2434	6 ABK11105	Abk11105 DNA enco
18	745	15.8	2431	6 ABK11104	Abk11104 DNA enco
19	736	15.6	2359	2 AAV07901	Aav07901 Cobra ven
20	736	15.6	2411	4 AAZ89683	Aaz89683 Human ADA
21	731.5	15.5	2050	2 AAV07895	Aav07895 Cobra ven
22	705	15.0	2240	6 ABK14952	Abk14952 Korean ad
23	702	14.9	1462	9 ADC38649	Adc38649 Human cDN
24	702	14.9	2187	7 ACA64909	Acac64909 Human dis
25	702	14.9	2209	2 AAT87427	Aat87427 Clone JS.
26	690	14.6	2280	2 AAV52604	Aav52604 Human met
27	677.5	14.4	2309	2 AAV07899	Aav07899 Cobra ven
28	674.5	14.3	2587	8 ADA27207	Ada27207 Murine ge
29	671.5	14.3	2297	2 AAV07896	Aav07896 Cobra ven
30	662	14.0	2288	2 AAV07898	Aav07898 Cobra ven
31	654.5	13.9	2335	2 AAV07897	Aav07897 Cobra ven
32	652.5	13.8	2475	8 ADA27195	Ada27195 Human MD8
33	648.5	13.8	3227	4 AAF30377	Aaf30377 Human ADA
34	648.5	13.8	3236	6 ABL64733	AbL64733 lung canc
35	648.5	13.8	3236	6 ABK40282	Abk40282 cDNA enco
36	648.5	13.8	3236	6 ABK83627	Abk83627 Human cDN
37	648.5	13.8	3236	9 ADC78888	Adc78888 Human PRO
38	648.5	13.8	3236	9 ADD18581	Add18581 Human dis
39	635	13.5	1851	5 AAC85063	Aac85063 E. carina
40	633	13.4	1863	7 ABX93715	Abx93715 Human thr
41	533	13.4	1863	7 ABX93715	Abx93715 Human thr
42	589.5	12.5	2029	3 AAZ89683	Aaz89683 Human ADA
43	589.5	12.5	2029	3 AAZ89683	Aaz89683 Human ADA
44	551.5	11.7	13715	6 ABO61152	Abog61152 MUC5B par
45	551	11.7	463	4 AAS41463	Aas41463 cDNA enco

ALIGNMENTS

RESULT 1

AAZ89683
ID AAZ89683 standard; DNA; 2325 BP.

XX AAZ89683;

DT 28-JUN-2000 (first entry)

DE Human ADAM DNA #2.

XX ADAM protein; human; A disintegrin and metalloprotease; diagnosis;

KW drug development; intervertebral hernia; sciatica; pulmonary fibrosis;
KW diabetic nephropathy; hepatic fibrosis; glomerulitis; osteopetrosis; ds.

OS Homo sapiens.

XX WC200014227-AL.

XX 16-MAR-2000.

PF 02-SEP-1999; 99WO-JP004766.

PR 03-SEP-1999; 98JP-00250115.

XX (TAKE } TAKEDA CHEM IND LTD.

PI Yoshimura K, Hikichi Y, Nishimura A;

XX WPI; 2000-271056/23.

DR P-PSDB; AAW08055.

PT Novel protein belong to A disintegrin and metalloprotease family, with

Qy	501	AlahsncdlyThrIysTrpGlyTrpAsnIysValTrpIleAsnIaGluTrpValTrpIle	620
Db	1801	GCCAAATGGAACTAAGTGTGGCATAAACAAGGTTTGCAATTAATCAGAAATGTGTGGATAT	1860
Qy	621	GluIysAlaTyrIysSerThrAsnTrpSerSerIysTrpIysGlyHisAlaValTrpTrp	640
Db	1861	GAGAAAGCCCTACAAATCAACCAATGTCTCATCTAAGTSCAAAGACATGCTGTGTGAC	1920
Qy	641	HisGluLeuGlnTrpGlnTrpGluGluGlyTrpIleProProTrpTrpTrpTrpSerSer	660
Db	1921	CATGAGCTCCAGTGTCAATGTGAGGAAGAGTGCCTCCCGACTGCGATGACTCTCTCA	1980
Qy	661	ValValTrpHisTrpSerIleValValGlyValLeuTrpProMetAlaValIleTrpVal	680
Db	1981	GTGGTCTTCCACTCTCCATTGTGTTGGGTGCTGTTCCTCAATGGGGTCATTTTTGTG	2040
Qy	681	ValValAlaMetValIleArgHisGlnSerSerArgGluIysGlnIysIysTrpGlnArg	700
Db	2041	GTGGTTGCTATGGTAATCCGGCACAGAGTCCAGAGAAAGCAGAAAGATCATCAGAGG	2100
Qy	701	ProLeuSerThrThrGlyThrArgProHisIysGlnIysArgIysProGlnMetValIys	720
Db	2101	CCACTATCTACCCTGSCACAGGCCACAAACAGAAAGAGAAACCCACAGATGTTTAAG	2160
Qy	721	AlaValGlnProGlnGluMetSerGlnMetIysProHisValTyrTrpLeuProValGlu	740
Db	2161	SGTGTTCMAECCCAAGAGATGAGTCAGATCAAGCCCATGTGTATGATCTGCCAGTAGAA	2220
Qy	741	GlyAsnGluProProAlaSerTrpHisIysTrpThrAsnAlaLeuProProThrValTrp	760
Db	2221	GGCATAGAGCCCCAGGCTCTTTTTCATTAAGACAAACCCACTTCCCCCTACTGTTTC	2280
Qy	761	IysTrpAsnProMetSerThrProIysTrpSerAsnProIysAla	775
Db	2281	AAGGATTAATCCAAATGTCTACCTTAAGACTCAATCCRAAGCA	2325

RESULT 2

AAZ89726	
AAZ89726	standard; DNA; 2828 BP.
XX	
XX	AAZ89726;
XX	
XX	28-JUN-2000 (first entry)
XX	
XX	Human ADAM DNA #4.
XX	
XX	ADAM protein; human; A disintegrin and metalloprotease; diagnosis;
KW	drug development; intervertebral hernia; sciatica; pulmonary fibrosis;
KW	diabetic nephropathy; hepatic fibrosis; glomerulitis; osteopetrosis; ds.
XX	
XX	
OS	Homo sapiens.
XX	
XX	WO200014227-A1.
XX	
XX	16-MAR-2000.
XX	
XX	02-SEP-1999; 99WC-JP004766.
PF	
XX	
XX	03-SEP-1998; 98JP-00250115.
PR	
XX	(TAKE) TAKEDA CHEM IND LTD.
XX	
XX	Yoshimura K, Hikichi Y, Nishimura A;
PI	
XX	
DR	WPI; 2000-271056/23.
DR	P-PSDE; AAW90865.
XX	
XX	Novel protein belong to A disintegrin and metalloprotease family, with
PT	protease activity and extracellular matrix digesting enzyme activity, for
PT	gene diagnosis and developing drugs for treating e.g. sciatica and
PT	glomerulitis.
XX	
XX	Disclosure; Fig 3-4; 109pp; Japanese.

QY 261 TrpLysIleIysIleThrProAsnAlaSerTrpThrLeuGluAsnTrpSerLysTrpArg 280
DB 835 GATPAGATTAAGATAACCCCAAGTGAAGCTTACCTTGGAGAAATTTTCTAATGAGG 894
QY 281 GlySerValLeuSerArgArgLysArgHisTrpIleAlaGlnLeuIleThrAlaThrGlu 300
DB 895 GGGAGTGTCTCTCAAGAGAAAGCGTCATGATATTGCTCAGTTAATCACAGCAACAGAA 954
QY 301 LeuAlaGlyThrThrValGlyLeuAlaTrpMetSerThrMetTrpSerProTrpSerVal 320
DB 955 CTTGCTGGAAACGACTGGGCTCTTGCACTTATCTTACATGTTGTTCTCTTATCTCTT 1014
QY 321 GlyValValGlnTrpHisSerTrpAsnLeuLeuArgValAlaGlyThrMetAlaHisGlu 340
DB 1015 GSGCTGTTCAGGACACACGCGATATCTTCTTAGAGTTGCGAGGAATGGCACATGAA 1074
QY 341 MetGlyHisAsnTrpGlyMetTrpHisTrpTrpTrpLysTrpProSerThrIle 360
DB 1075 ATGGGCAACAACCTTGGAAATGTTTCATGCGACTATTCTTGCAGTGTCTCTTACAATA 1134
QY 361 TrpValMetTrpLysAlaLeuSerTrpTyrIleProThrTrpTrpSerSerTrpSerArg 380
DB 1135 TGTGTGATGGACAAACGACCTGAGCTTCTATATACCCACACAGACTTCAGTTCTCGACCGT 1194
QY 381 LeuSerTrpTrpLysTrpTrpGluTrpLysLeuSerAsnTrpLeuTrpAsnAlaProLeu 400
DB 1195 CTCAGCTATGACAAAGTTTTTGAAGATAAATATCAATGCTCTTTTAATGCTCCATGTG 1254
QY 401 ProThrTrpIleIleSerThrProIleTrpGlyAsnGlnLeuValGluMetGlyGluTrp 420
DB 1255 CCTACAGATATCATATCCACTCCAAATTTGTGGAAACAGTTGGTGGAATGGGAGAGGAC 1314
QY 421 TrpTrpTrpGlyThrSerGluGluTrpTrpAsnIleTrpTrpTrpAlaLysTrpTrpLys 440
DB 1315 TGTGATGCTGGACATCTCGAGGAATGTACCAATATTTGCTGTGATGCTTAAGACATGTAA 1374
QY 441 IleLysAlaThrTrpGlnTrpAlaLeuGlyGluTrpTrpGluLysTrpGlnTrpLysLys 460
DB 1375 ATCAAGCAACCTTTCAATGTGCAITAGGAGAAATGTTGTGMAAAATGCCAATTTAABAAG 1434
QY 461 AlaGlyMetValTrpArgProAlaLysTrpGluTrpTrpLeuProGluMetTrpAsnGly 480
DB 1435 GCTGGGATGGTGTGCAGACAGCAAGATGATGCGACCTGCTGCAATGTGTAATGCT 1494
QY 481 LysSerGlyAsnTrpProTrpTrpArgTrpGlnValAsnGlyTrpProTrpHisGly 500
DB 1495 AATCTGTAATGCTTCGATGATAGATCCAAAGTCATGGCTTCCTTGCCTGCAACGGG 1554
QY 501 LysGlyHisTrpLeuMetGlyThrTrpProThrLeuGlnGluGlnTrpThrGluLeuTrp 520
DB 1555 AAGGGCCACTGCTTGTATGGGACATGCCCCACACTGACAGAGCAGTGCAGAGCTGTGG 1614
QY 521 GlyProGlyThrGluValAlaTrpLysSerTrpTyrAsnArgAsnGluGlyGlySerLys 540
DB 1615 GGACCAAGAACTGAGGTGACAGTAAGTCATGTACACAGGAATGAAGTGGGTCAAG 1674
QY 541 TyrGlyTyrTrpArgArgValTrpTrpThrLeuIleProTrpLysAlaAsnTrpThrMet 560
DB 1675 TACGGGTACTGCGAGAGTGGATGACACACTCATCTCCCTGCAAGCAATGATACCATG 1734
QY 561 TrpGlyLysLeuTrpTrpGlnGlyGlySerTrpAsnLeuProTrpLysGlyArgIleVal 580
DB 1735 TGTGGGAAGTTCCTGCTCAGGTGGGTGGGATAATTTGCCCTGGAAAGACGATAGTG 1794
QY 581 ThrTrpLeuThrTrpLysThrTrpTrpProGluTrpTrpThrSerGlnGluIleGlyMetVal 600
DB 1795 ACTTTCCTACATGTAAACATTTGATCCTGAAGACACAAAGTCACAGAAATAGGCGATGGT 1854
QY 601 AlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpIleAsnAlaGlnTrpValTrpIle 620
DB 1855 GCCAATGNACTAAGTGGCGCATACAAAGTTTGCATTAATGCAAGATGTGGATATT 1914
QY 621 GluLysAlaTyrLysSerThrAsnTrpSerSerLysTrpLysGlyHisAlaValTrpTrp 640

DB 1915 GAGAAAGCCCTACAAATACCAATTCCTCACTAAGTGAAGGACATGCTGTGTGAC 1974
QY 641 HisGluLeuGlnTrpGlnTrpGluGluGlyTrpIleProTrpTrpTrpTrpTrpSer 660
DB 1975 CATGAGCTCCAGTGTCAATGTGAGGAAGATGATCCCTCCCGACTGCGATGACTCTCA 2034
QY 661 ValValTrpHisTrpSerIleValValGlyValLeuTrpProMetAlaValIleTrpVal 680
DB 2035 GTGGTCTTCCACTTCTCCATGTGTGGTGGGTGCTGTTCCCAATGGCGGTCAATTTTGTG 2094
QY 681 ValValAlaMetValIleArgHisGlnSerSerArgGluLysGlnLysLysTrpGlnArg 700
DB 2095 GTGGTGTCTATGTATATCCGACACAGAGCTCCAGAGAAAGCAGAAAGATCATCAGG 2154
QY 701 ProLeuSerThrThrGlyThrArgProHisLysGlnLysArgLysProGlnMetValLys 720
DB 2155 CCCTATCTTACCACTGGCAGCCACACAAACAGAGAGGAAACCCAGATGGTAAG 2214
QY 721 AlaValGlnProGlnGluMetSerGlnMetLysProHisValTrpTrpLeuProValGlu 740
DB 2215 GTGTGTTCAACCCCAAGAGATGATGATGAGGCCCATGTGTATGATCTCCAGTAGAA 2274
QY 741 GlyAsnGluProProAlaSerTrpHisLysTrpThrAsnAlaLeuProTrpValTrp 760
DB 2275 GGCATGATGCCCCCAGCCCTCTTTTCATTAAGACACAAACGACCTTCCCTACTGTTTC 2334
QY 761 LysTrpAsnProMetSerThrProLysTrpSerAsnProLysAla 775
DB 2335 AAGGATATCCAAATGTCTACCTAAGGACTCAATCCAAAGCA 2379
RESULT 3
ABR09173
ID ABA09173 standard; cDNA; 3179 BP.
XX ABA09173;
XX 11-JAN-2002 (first entry)
DE Human eMDC II protein homologue-encoding cDNA, SEQ ID NO:949.
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytotatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnery; antiulcer; ss.
OS Homo sapiens.
XX
XX WO200157188-A2.
XX
PD 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US003800.
XX
XX 03-FEB-2000; 2000US-00496914.
XX 27-APR-2000; 2000US-00560875.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-457740/49.
XX P-PSDB; ABB11929.
XX

PT Human proteins and DNA encoding sequences useful for preventing, treating
 PT or ameliorating a medical condition in a mammalian subject e.g. arthritis
 PT and cancer.

XX Claim 1; Page 812-813; 1963pp; English.

XX Sequences ABB10981-ABBI2330 represent 1350 novel human polypeptides, and
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
 CC invention also relates to vectors and recombinant host cells comprising a
 CC nucleotide of the invention, methods of producing the novel polypeptides,
 CC antibodies against the polypeptides, methods of detecting the nucleotides
 CC or polypeptides in a sample, and methods of identifying compounds which
 CC bind to polypeptides of the invention. Although novel, many of the
 CC polypeptides of the invention have homology to known proteins, thereby
 CC giving an insight into their probable biological activities, and hence
 CC potential therapeutic applications. The polypeptides of the invention may
 CC have various activities, including cytokine, cell proliferation or cell
 CC differentiation activities; stem cell growth factor activity;
 CC haematopoiesis regulatory activity; tissue growth activity;
 CC immunomodulatory activity; activin- or inhibin-related activities;
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
 CC thrombolytic activities; receptor or ligand activities; or may be
 CC involved in oncogenesis, cancer cell proliferation or metastasis.
 CC Depending on their biological activities, polypeptides and nucleotides of
 CC the invention are useful for preventing, treating or ameliorating medical
 CC conditions, e.g., by protein or gene therapy. Such conditions include
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
 CC vascular growth. Polypeptides involved with tissue regeneration and
 CC repair (or nucleic acids encoding them) may be used to promote wound
 CC healing (e.g., of burns, incisions and ulcers), while those with
 CC immunomodulatory activities may be used in the treatment of viral,
 CC bacterial and fungal infections in addition to immune disorders.
 CC Polypeptides with growth factor activity may be used in cell cultures to
 CC promote cell growth. For example, such polypeptides may be used to
 CC manipulate stem cells in culture to give rise to neuroepithelial cells
 CC that can be used to augment or replace cells damaged by illness,
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides
 CC may also be used in the diagnosis of the above conditions, and in drug
 CC screening techniques. The present sequence represents a cDNA encoding a
 CC novel human polypeptide of the invention

XX SQ Sequence 3179 BP; 1042 A; 581 C; 692 G; 864 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.24e-170	Length:	3179
Score:	3128.00	Matches:	654
Percent Similarity:	88.13%	Conservative:	29
Best Local Similarity:	84.39%	Mismatches:	92
Query Match:	66.38%	Indels:	0
DB:	4	Gaps:	0

US-10-726-148A-15 (1-775) x ABA09173 (1-3179)

Qy	1	MetLeuGlnGlyLeuLeuProValSerLeuLeuSerValAlaValSerAlaIleLys	20
Db	48	ATGTGTGAAGGTCCTCGCCAGTCAGTCTCTCTCTCTCTGTGTGAGTAAGTGCTATRAAA	107
Qy	21	GluLeuProGlyValLysLysTyrGluValValTyrProIleArgLeuHisProLeuHis	40
Db	108	GAACCTCCCTGGGGTGAAGAAGTATGAAGTGTATCTCTATTAAGTATCACTCCATGAT	167
Qy	41	LysArgGluAlaLysGluProGlnGlnGlnGlnGlnTyrGluThrGluLeuLysTyrLys	60
Db	168	AAAAAGAGAGGCCAAAGACCCAGAGCAACAGCAAAATTTGAACCTGAATTAAAGTATAAA	227
Qy	61	MetThrLeuSerGlyLysIleAlaValLeuTyrLeuLysLysAsnLysAsnLeuLeuAla	80
Db	228	ATGCAATTAATGAAAAAATTCAGTGTCTTTATTGAAAAAACAAGAACCTCCTTGCA	287
Qy	81	ProGlyTyrThrGluThrTyrTyrAsnSerThrGlyLysGluIleThrThrSerProGln	100

Db	288	CCAGGCTACACGGAAACATATTATAATCCACTGGAAGAGGAGATCACCACAAAGCCCAAA	347
Qy	101	IleMetTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyr	120
Db	348	ATTATGGATGATTGTTATATCAAGGACATATCTTATGAAAGAGTTTCTGACGCTAGC	407
Qy	121	IleSerThrTyrArgGlyLeuArgGlyTyrTyrSerGlnGlyTyrGlnArgTyrTrpIle	140
Db	408	ATCAGCACATAGGGGCTTAAGGGGCTACTTCACTCAGGGGGATCAAAAGACTATTATT	467
Qy	141	GluProLeuSerProIleHisArgTyrGlnGlnHisAlaLeuTyrLysTyrAsnPro	160
Db	468	GAACCTTTAACCCCATACATCGGATCGACAGGAGCATGCTCTTCAAGTATTAACCT	527
Qy	161	TrpGluLysAsnTyrTyrSerThrTyrGlyMetTyrGlyValLeuTyrAlaHisTyrLeu	180
Db	528	GATGAAAGAAATTAATGACAGCACCTGTGGATGGATGGTGTGTGTGGCCCAAGATTG	587
Qy	181	GlnGlnAsnIleAlaLeuProAlaThrLysLeuValLysLeuLysTyrArgLysValGln	200
Db	588	CAGCAGAACATTTGCCCTTACCTGCCCAACAACTAGTAAATTTGAAAGACAGGAAGTTT	647
Qy	201	GluHisGluLysTyrIleGlyTyrTyrLeuValLeuTyrAsnGlyGluTyrLysArgTyr	220
Db	648	GAACATGGAATACATAGATATATTTTGGTCTGGATATATGGTGGATTTAAAGGTAC	707
Qy	221	AsnGluAsnGlnTyrGluIleArgLysArgValTyrGluMetAlaAsnTyrValAsnMet	240
Db	708	AATCAGAAATCAAGATGAGATCAGAAAGAGGGTATTTCAGATGGCTAAATTATCTCAACATG	767
Qy	241	LeuTyrLysLysLeuAsnThrHisValAlaLeuValGlyMetGluIleTyrThrTyrLys	260
Db	768	CTTTATAAAAAGCTCAATPACTCATGTGGCTTAGTGGTATGTAATGGAATCTGGACTGACAA	827
Qy	261	TrpLysIleLysIleThrProAsnAlaSerTyrThrLeuGluAsnTyrSerLysTyrArg	280
Db	828	GATTAAGATTAAGATTAACCCCAATGCAAGCTTCACTTGGAGAAATTTCTTAATGGAGG	887
Qy	281	GlySerValLeuSerArgArgLysArgHisTrpIleAlaGlnLeuIleThrAlaThrGlu	300
Db	888	GGGAGTGTCTCTCAAGAGAAAGCGTCAATGATATGGCTAGTTAATCAGCAACAGAA	947
Qy	301	LeuAlaGlyThrThrValGlyLeuAlaTyrMetSerThrMetTyrSerProTyrSerVal	320
Db	948	CTTGTGGAACGACTGTGGGTCTTGCAATTATGTCTCAAAATGTGTCTCTCTTATCTGT	1007
Qy	321	GlyValValGlnTyrHisSerTyrAsnLeuArgValAlaGlyThrMetAlaHisGlu	340
Db	1008	GGCGTGTTCAGGACCAAGCGATATCTTCTAGAGTTGCGAGCAATGGCAATGAA	1067
Qy	341	MetGlyHisAsnTyrGlyMetTyrHisTyrTyrTyrSerTyrLysTyrProSerThrIle	360
Db	1068	ATGGGCCACAACTTTGGAATGTTTCAACGACTATTCTTGCAGTGTCTCTTCTTCAATA	1127
Qy	361	TrpValMetTyrLysAlaLeuSerTyrTyrIleProThrTyrTyrSerTyrSerArg	380
Db	1128	TGTGTGATGGCAAGCAAGCACTCAGCTTCTATATACCCACAGACTTTCAGTCTCGACCGT	1187
Qy	381	LeuSerTyrTyrLysTyrTyrLysLeuSerAsnTyrLeuTyrAsnAlaProLeu	400
Db	1198	CTCAGCTATGACAGATTTTGAAGATTAATATCAATTCCTCTTTAATGCTCCATTG	1247
Qy	401	ProThrTrpIleLeuSerThrProIleTyrGlyAsnGlnLeuValGluMetGlyGluTyr	420
Db	1248	CCTACAGATATCATATCCACTCCCAATTTGTGGGAACCAAGTTGGTGGAAATGGAGAGAC	1307
Qy	421	TrpTyrTyrGlyThrSerGluGluTyrThrAsnIleTyrTyrTyrAlaLysThrTyrLys	440
Db	1308	TGTGATTGTGGGACATCTGAGGAATGTACCAATATTCCTGTGATGCTTAAGACATGTAAA	1367
Qy	441	IleLysAlaThrTyrGlnTyrAlaLeuGlyGluTyrTyrGluLysTyrGlnTyrLys	460

Db	1368	ATCAAGACGACATTTTCAATGTGCTATTAGAGAAATGTTGTGAAATAATGCCAAATTTAAAGAAG	1427
Qy	461	AlaGlyMetValTrrArgProAlaIbysTrpGluTrpTrpLeuProGluMetTrpAsnGly	480
Db	1428	GCTGGGATGGTGTGCAGACACGACAAAGATGAGTGGACCTGCCTGAAATGTTGAATGGT	1487
Qy	481	LysSerGlyAsnTrpProTrpTrpArgTrpGlnValAsnGlyTrpProTrpHisHisGly	500
Db	1488	AAATCTGTGTAATGTCCTTGATGATAGAGATCCAAAGTCATGCTTCCCTTGCCTACACGGG	1547
Qy	501	LysGlyHisTrpLeuMetGlyThrTrpProThrLeuGlnGluGlnTrpThrGluLeuTrp	520
Db	1548	AAGGGCCACTGCTTGATGGGGACATGCCCHACATGACGGAGCAGTGCACAGAGCTGTGG	1607
Qy	521	GlyProGlyThrGluValAlaIrrIlySerTrpTyrAsnArgAsnGluGlyGlySerIlys	540
Db	1608	GGACGAGAACTGAGGTTCAGATAAGTCATGTTTACACAGGAATGAAGGTGGGTCAAG	1667
Qy	541	TyrGlyTyrTrpArgArgVal.TrrTrpThrLeuLeuProTrpLysAlaAsnTrpThrMet	560
Db	1668	TACGGGTACTGTGCAGAGTGGATGCACACATCTCATTCCTCCAAAGCAATGATACCATG	1727
Qy	561	TrpGlyLysLeuLeuTrpGlnGlyGlySerTrpAsnLeuProTrpLysGlyValGluVal	580
Db	1728	TGTGGAGATGTTCTGTCAGGTGGGTGGATATTGTCCTCGAAGACAGCGATAGT	1787
Qy	581	ThrTrpLeuThrTrpLysThrTrpTrpProGluTrpThrSerGlnGluIrrGlyMetVal	600
Db	1788	ACTTTCCTGCATCTAAACATTTTGATCCTCGAAGACACAGCTCAAGAAATAGGCATGGT	1847
Qy	601	AlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpIleAsnAlaGluTrpValTrpIle	620
Db	1848	GCCAAATGGAACTAAGTGGCGGATAACAGGTTTGCATTATATGAGAAATGTGTGGATATT	1907
Qy	621	GluLysAlaTrpLysSerThrAsnTrpSerSerLysTrpLysGlyHisAlaValTrpTrp	640
Db	1908	GAGAAAGCCATCAAAATCAACCAATGTCTCTCTAGTGCAAGAGCAATGCTGTGTGTGAC	1967
Qy	641	HisGluLeuGlnTrpGlnTrpGluGluGlyTrpIleProProThrTrpTrpTrpSerSer	660
Db	1968	CATGAGTCCAGTCTCAATGTGAGGAAGGATGGATCCTCCGACTCGCATGACTCTCCA	2027
Qy	661	ValValTrpHisTrpSerIleValValGlyValLeuTrpProMetAlaValIleTrpVal	680
Db	2028	GTGGTCTTTCACACTCTCATGTGTTGGGTGTCTGTTCCTCAATGGCGGTCAITTTTGTG	2087
Qy	681	ValValAlaMetValIleArgHisGlnSerSerArgGluLysGlnLysLysTrpGluArg	703
Db	2088	GTGGTGTCTATGGTAATCCGGCACCGAGCTCCAGAGAAAGCAGAGAAGATCAGAGG	2147
Qy	701	ProLeuSerThrThrGlyThrArgProHisLysGlnLysArgLysProGlnMetValLys	720
Db	2148	CCACTATCTACCACTGGCACCCAGGCCACACAAACAGAAAGAGAAACCCACAGATGGTAAAG	2207
Qy	721	AlaValGlnProGlnGluMetSerGlnMetLysProHisValTyrTrpLeuProValGlu	740
Db	2208	GCTGTTCAACCCCAAGAGATGAGTCAGATGAAGCCCCCATGTGTGTGATCTGCCAGTAGAA	2267
Qy	741	GlyAsnGluProProAlaSerTrpHisLysTrpThrAsnAlaLeuProProThrValTrp	760
Db	2268	GGCAATCAGCCCCCAGCCTCTTTTCAAAAGACACAAACGACATCTCCCTACTGTGTTTC	2327
Qy	761	LysTrpAsnProMetSerThrTrpLysTrpSerAsnProLysAla	775
Db	2328	AAGGATATCCAAATSTCTACACTTAAGGACTCCAAATCCAAAGCA	2372

RESULT 4

RESULT 4
AAZ89673

ID: AAZ89673 standard; DNA; 1620 BP.

XX

62

XX DE

ET 28-JUN-2000 (first entry)

Db	1448	AAATCTGGTAAATGCTCTGATGATAGATCCAAAGTCAATGCTTCCCTTGCCTACACGGG	1507
Qy	501	LysGlyHisTrpLeuMetGlyThrTrpProThrLeuGlnGluGlnTrpThrGluLeuTrp	520
Db	1508	AAAGGGCACTGCTTGTATGGGGACATGCCACACTGCAGGAGCAGTGCACAGAGCTGTGG	1567
Qy	521	GlyProGly	523
Db	1568	GGACACAGT	1576
RESULT B			
AAH33351	AAH33351 standard; cDNA; 2047 BP.		
XX	AC	AAH33351;	
XX	DC	03-SEP-2001 (first entry)	
XX	DE	Human colon cancer antigen encoding cDNA SEQ ID NO:407.	
XX	KW	Human; colon cancer; colon cancer antigen; diagnosis; detection;	
XX	KW	colorectal carcinoma; ss.	
XX	OS	Homo sapiens.	
XX	XX	W0200122920-A2.	
XX	XX	05-APR-2001.	
XX	XX	28-SEP-2000; 2000MO-US026524.	
XX	XX	29-SEP-1999; 99US-0157137P.	
XX	XX	03-NOV-1999; 99US-0163280P.	
XX	XX	(HUMA-) HUMAN GENOME SCI INC.	
XX	PI	Ruben SM, Barash SC, Birse CE, Rosen CA;	
XX	XX	WPI; 2001-235357/24.	
XX	DR	P-PSDB; AAG73920.	
XX	PT	Nucleic acids encoding 4277 human colon cancer-associated polypeptides,	
XX	PT	useful for preventing, diagnosing and/or treating colorectal cancers.	
XX	PS	Claim 1; Page 2512; 9803pp; English.	
XX	CC	AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon	
XX	CC	cancer-associated nucleic acid molecules (N) and proteins (P), where the	
XX	CC	proteins are collectively known as colon cancer antigens. The colon	
XX	CC	cancer antigens have cytostatic activity and can be used in gene therapy	
XX	CC	and vaccine production. N and P may be used in the prevention, diagnosis	
XX	CC	and treatment of diseases associated with inappropriate P expression. For	
XX	CC	example, N and P may be used to treat disorders associated with decreased	
XX	CC	expression by rectifying mutations or deletions in a patient's genome	
XX	CC	that affect the activity of P by expressing inactive proteins or to	
XX	CC	supplement the patient's own production of P. Additionally, N may be used	
XX	CC	to produce the colon cancer-associated Ps, by inserting the nucleic acids	
XX	CC	into a host cell and culturing the cell to express the proteins. N and P	
XX	CC	can be used in the prevention, diagnosis and treatment of colorectal	
XX	CC	carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent	
XX	CC	sequences used in the exemplification of the present invention. N.B.	
XX	CC	Pages 666 to 682 and page 7053 of the sequence listing were missing at	
XX	CC	time of publication, meaning no sequences are present for SEQ ID NO:1027	
XX	CC	to 1052, 7921 and 7922	
XX	SQ	Sequence 2047 BP; 670 A; 372 C; 443 G; 562 T; 0 U; 0 Other;	
Alignment Scores:			
Pred. No.:	1.27e-83	Length:	2047
Score:	1610.00	Matches:	345
Percent Similarity:	83.72%	Conservative:	20
Best Local Similarity:	79.13%	Mismatches:	71
Query Match:	34.17%	Indels:	0

Db 1031 GTGGTGGTTCATGTAATCGGACCAAGCTCCAGAAAGCAGAAAGATCAG 1090
 QY 700 ArgProLeuSerTrpThrGlyThrArgProHisGlyLysGlnLysArgLysProGlnMetVal 719
 Db 1091 AGGCCACTATCTACCACTGGCACCAGGCCACACAAACAGAAAGAGAAACCCAGATGGTA 1150
 QY 720 LysAlaValGlnProGlnGluMetSerGlnMetLysProHisValTrpLeuProVal 739
 Db 1151 AAGGCTGTTCACCCCAAGAGATGAGTCAGATGAAGCCCCCATGGTAAGATCTGCCAGTA 1210
 QY 740 GluGlyAsnGluProProAlaSerTrpHisLysTrpThrAsnAlaLeuProProThrVal 759
 Db 1211 GAAGGCAATGAGCCCCAGGCTCTTTTCAATAAGACACAAAGCAGCTTCCCCCTACTGTT 1270
 QY 760 TrpLysTrpAsnProMetSerTrpProLysTrpSerAsnProLysAla 775
 Db 1271 TTCAAGGATAATCAATGTCTACACCTAAGGACTCAAAATCCAAAAGCA 1318

RESULT 9

AAC91179
 ID AAC91179 standard; DNA; 2047 BP.

AC AAC91179;

XX 20-MAR-2001 (first entry)

DT ADAM gene #3.

XX ADAM; disintegrin; metalloprotease; food additive; breast cancer;
 XX ovarian; ss.

OS Homo sapiens.

XX W0200073323-A2.

XX 07-DEC-2000.

XX 25-MAY-2000; 2000WO-US014308.

XX 27-MAY-1999; 99US-0135388P.

XX 29-JUL-1999; 99US-0142930P.

XX 28-JAN-2000; 2000US-0178717P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Wei P, Ni J, Hastings GA, Shi Y;

XX WPI; 2001-016507/02.

XX Seven nucleic acid molecules encoding ADAM polypeptides containing a
 PT disintegrin and metalloprotease domain, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases.

XX Claim 1; Page 267; 287pp; English.

XX The present invention relates to seven members of the ADAMs (proteins
 CC which contain A Disintegrin And Metalloprotease domain) protein family.
 CC The ADAMs proteins and DNA may be used to treat disease, as a food
 CC additive or preservative, for chromosome identification, as probes for
 CC diagnosing a disorder related to the female reproductive system,
 CC particularly breast and/or ovary cancer. They are also useful in the gene
 CC therapy of breast and ovarian cancer

XX Sequence 2047 BP; 670 A; 372 C; 443 G; 562 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1-27e-83 Length: 2047
 Score: 1610.00 Matches: 345
 Percent Similarity: 83.72% Conservative: 20
 Best Local Similarity: 79.13% Mismatches: 71
 Query Match: 34.17% Indels: 0

DB: 4 Gaps: 0
 US-10-726-148a-15 (1-775) x AAC91179 (1-2047)
 QY 340 GluMetGlyHisAsnTrpGlyMetTrpHisTrpTrpTyrSerTrpLysTrpProSerThr 359
 Db 11 GAAATGGGCCACAACTTTGGATGTTTCATGACGACTATTCTTCGAAAGTGTCTTCTACA 70
 QY 360 IleTrpValMetTrpLysAlaLeuSerTrpTyrLleProThrTrpTrpSerSerTrpSer 379
 Db 71 ATATGTGTGATGAGCAAGCACTGAGCTTCTATATACCCACAGACTTCAGTTCCTGCAGC 130
 QY 380 ArgLeuSerTrpTrpLysTrpTrpGluTrpLysLeuSerAsnTrpLeuTrpAsnAlaPro 399
 Db 131 CGTCTCAGCTATGCAAGTTTTTGAAGATAAATATCAAAATTCCTCTTTAATGCTCCA 190
 QY 400 LeuProThrTrpLleLeSerThrProIleTrpGlyAsnGlnLeuValGluMetGlyGlu 419
 Db 191 TTGCCTACAGATATCATCCACTCCAATTGTGGGAACCAAGTGGTGGAAATGGGAGAG 250
 QY 420 TrpTrpTrpTrpGlyThrSerGluGluTrpThrAsnIleTrpTrpTrpAlaLysThrTrp 439
 Db 251 GACTGTGATTTGGGACATCTGAGGAATGTACCAATATTTGCTGTGATGCTAAACACATGT 310
 QY 440 LysIleLysAlaThrTrpGlnTrpAlaLeuGlyGluTrpTrpGluLysTrpGlnTrpLys 459
 Db 311 AAAATCAAGCACTTTTCAATGTGCATTAGGAGATGTTGTGAAAATGCCAATTTAAA 370
 QY 460 LysAlaGlyMetValTrpArgProAlaLysTrpGluTrpTrpLeuProGluMetTrpAsn 479
 Db 371 AAGGCTGGGATGGTGTGCAGACCAGCAAAAGATGAGTGGCAGCTGCCCTGAAATGTGAAT 430
 QY 480 GlyLysSerGlyAsnTrpProTrpTrpArgTrpGlnValLeuGlyTrpProTrpHisHis 499
 Db 431 GGTAAATCTGTGTAATTTGCTGTGATGATAGATTCCAAGTCATGGCTTCCCTTGCCATCAC 490
 QY 500 GlyLysGlyHisTrpLeuMetGlyThrTrpProThrLeuGlnGluTrpThrGluLeu 519
 Db 491 GGGAGGGGCCACTCTTGTGATGGGCACATGCCACACTGCAGAGGAGCTGCACAGAGCTG 550
 QY 520 TrpGlyProGlyThrGluValAlaTrpLysSerTrpTrpAsnArgAsnGluGlyGlySer 539
 Db 551 TGGGACCCAGCAACTGAGGTTGCAGATAAGTTCATGTTACAAACAGGAATGAAGGTGGTCA 610
 QY 540 LysTrpGlyTrpTrpArgArgValTrpTrpThrLeuIleProTrpLysAlaAsnTrpThr 559
 Db 611 AAGTACGGGTACTGTCGAGAGTGATGACACACTCATTTCCCTGCAAGCAAAATGATATACC 670
 QY 560 MetTrpGlyLysLeuTrpTrpGlnGlySerTrpAsnLeuProTrpLysGlyArgIle 579
 Db 671 ATGTGTGGGAAGTTGTTCTGTCAAGTGGTGGTGGATAAATTTGCCCTGGAAGGACGGATA 730
 QY 580 ValThrTrpLeuThrTrpLysThrTrpProGluTrpThrSerGlnGluIleGlyMet 599
 Db 731 GTGACTTTCTCTGACATGTAAACATTTGATCCTCAAGACACACAAAGTCAAGAAATAGGCATG 790
 QY 600 ValAlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpIleAsnAlaGluTrpValTrp 619
 Db 791 GTGGCCCAATGAACTAAGTGTGGCGATACAAAGTTTGCATTATGAGAGATGTGTGAT 850
 QY 620 IleGluLysAlaTrpLysSerThrAsnTrpSerSerLysTrpLysGlyHisAlaValTrp 639
 Db 851 ATTGAGAAAGCCTACAAATCAACCAATTTGCTCATCTAAGTCAAAAGGACATGCTGTGTGT 910
 QY 640 TrpHisGluLeuGlnTrpGlnTrpGluGlyTrpIleProProTrpTrpTrpTrpSer 659
 Db 911 GACCATGAGCTCCAGTGTCAATGTGAGAGAGATGAGTCCCTCCCGAGTGCATGACATCC 970
 QY 660 SerValValTrpHisTrpSerIleValValGlyValLeuTrpProMetAlaValIleTrp 679
 Db 971 TCAGTGTCTCTCCACTTCTCCATTTGTGGGTGCTGTTCCTCAATGGCGGTCAATTTT 1030
 QY 680 ValValValAlaMetValIleArgHisGlnSerSerArgGluLysGlnLysLysTrpGln 699

DB 1031 GTGGTGGTGGTATGGTAATCCGACACAGGCTCCAGAGAAAGAGAGAGATCAG 1090
QY 700 ArgProLeuSerThrGlyThrArgProHisLysGlnLysArgLysProGlnMetVal 719
DB 1091 AGGCCACTATCTACCACTGGCCAGGCGCCACACAAACAGAGAGAAACCCAGATGTA 1150
QY 720 LysAlaValGlnProGlnGluMetSerGlnMetLysProHisValTyrTrpLeuProVal 739
DB 1151 AAGGCTGTTCAACCCCAAGAGGTGAGTCAGATGAAGGCCCAATGTGTATGATCTGCCAGTA 1210
QY 740 GluGlyAsnGluProProAlaSerTrpHisLysTrpThrAsnAlaLeuProProThrVal 759
DB 1211 GAAGGCAATGAGCCGCCAGCCCTCTTTTCATTAAGACACAAAGC&CCTTCCCTACTGTT 1270
QY 760 TrpLysTrpAsnProMetSerThrProLysTrpSerLysProLysAla 775
DB 1271 TTCAAGGATAATCCATGTCTACACCTAAGGACTCAAAATCCAAAGCA 1318
RESULT 10
AAF33052
ID AAF33052 standard; cDNA; 2047 BP.
XX
AC AAF33052;
XX
DT 23-MAR-2001 (first entry)
XX
DE Human secreted protein gene 16 SEQ ID NO:26.
XX
KW Human; secreted protein; diagnosis; cytostatic; antirheumatic;
KW antiarthritic; dermatological; cardiac; antinflammatory; anti-ulcer;
KW gastrointestinal; solid tumour; rheumatoid arthritis; psoriasis;
KW diabetic retinopathy; myocardial angiogenesis; Crohn's disease; ulcer;
KW ss.
XX
OS Homo sapiens.
XX
PN W0200077237-A1.
XX
PD 21-DEC-2000.
XX
PF 01-JUN-2000; 2000MO-US014928.
XX
PR 11-JUN-1999; 99US-0138633P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/; ROSEN C A.
XX
PI Rosen CA, Ruben SM, Komatsoulis GA;
XX
XX WPI: 2001-071280/08.
DR P-PSDB; AAB64681.
XX
PT Nucleic acids encoding 49 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating diseases such as tumors,
PT rheumatoid arthritis, psoriasis and diabetic retinopathy.
PS
PS Claim 1; Page 439-440; 520pp; English.
XX
CC The polynucleotide sequences given in AAF33037 to AAF33085 encode the
CC human secreted proteins given in AAB64666 to AAB64714. AAB64715 to
CC AAB64771 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are given in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissues
CC and cells the genes are expressed in. Examples of activities include:
CC cytostatic; antirheumatic; antiarthritic; dermatological; cardiac;
CC antinflammatory; gastrointestinal; and anti-ulcer. The polynucleotides
CC and polypeptides can be used in the prevention, treatment and diagnosis
CC of diseases associated with inappropriate polypeptide expression.
CC Disorders that may be treated or prevented include solid tumours,
CC rheumatoid arthritis, psoriasis, diabetic retinopathy, myocardial
CC angiogenesis, Crohn's disease and ulcers. The polynucleotides and their
CC complementary sequences may also be used as DNA probes in diagnostic

assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate
the presence of similar nucleic acid sequences in samples, and therefore
which patients may be in need of restorative therapy. The polypeptides
may also be used as antigens in the production of antibodies against the
polypeptide and in assays to identify modulators (agonists and
antagonists) of polypeptide expression and activity. The anti-polypeptide
antibodies and antagonists may also be used to down regulate expression
and activity. AAF33028 to AAF33036 and AAB64665 represent sequences used
in the exemplification of the present invention

XX Sequence 2047 BP; 670 A; 372 C; 443 G; 562 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1,27e-83 Length: 2047
Score: 1610.00 Matches: 345
Percent Similarity: 83.72% Conservative: 26
Best Local Similarity: 79.13% Mismatches: 71
Query Match: 34.17% Indels: 0
DB: 4 Gaps: 0

US-10-726-148A-15 (1-775) x AAF33052 (1-2047)

QY 340 GluMetGlyHisAsnTrpGlyMetTrpHisTrpTrpLysTrpLysTrpSerThr 359
DB 11 GAAATGGGCCACAACCTTTGGAATGTTTCATGAGCACTATCTTGCAGTGTCTTCTACA 70
QY 360 IleTrpValMetTrpLysAlaLeuSerTrpTrpLysTrpTrpLysTrpSerThr 379
DB 71 ATATGTGTGATGGACAAAGCACTGAGCTTCTATATACCCACAGACTTCAGTCTCTGCAGC 130
QY 380 ArgLeuSerTrpTrpLysTrpGluTrpLysLeuSerAsnTrpLeuTrpAsnAlaPro 399
DB 131 CGTCTCAGTATGACAGATTTTTCAGAGATAATATTCAMATGCTCTCTTAAATGCTCCA 190
QY 400 LeuProThrTrpLysLysSerTrpProLysTrpGlyAsnGlnLeuValGluMetGlyGlu 419
DB 191 TTGCTACAGATATCATATCCACTCCAATTTGTGGGAACCACTGTGGTGAATGGAGAG 250
QY 420 TrpTrpTrpGlyThrSerGluGluTrpThrAsnLysTrpTrpTrpAlaLysThrTrp 439
DB 251 GACTGTGATTTGGGACATCTGAGGAAATGTACCAATATTTGCTGTGATGCTAAGACATGT 310
QY 440 LysLysLysAlaThrTrpGlnTrpAlaLeuGlyGluTrpTrpGluLysTrpGlnTrpLys 459
DB 311 ABAATCAAGCAACTTTTCAATGTGCATTAGAGAAATGTTGAAAAATGCCAATTAAAA 370
QY 460 LysAlaGlyMetValTrpArgProAlaLysTrpGluTrpTrpLeuProGluMetTrpAsn 479
DB 371 AAGGCTGGGATGGTGTGCAGACCAAGATGATGCGACCTGCCTGAAATGTGTAAT 430
QY 480 GlyLysSerGlyAsnTrpProTrpTrpArgTrpGlnValAsnGlyTrpProTrpHisHis 499
DB 431 GGTAAATCTGGTAATTTGCTGATGATAGATCCAAAGTCATGGCTTCCCTTGCCTATCAC 490
QY 500 GlyLysGlyHisTrpLeuMetGlyThrTrpProThrLeuGlnGluTrpTrpGluLeu 519
DB 491 GGGAAAGGCGCACTGCTTCAATGGGACATGCCCACTGCAGGAGCAGTGCACAGAGCTG 550
QY 520 TrpGlyProGlyThrGluValAlaTrpLysSerTrpTrpAsnArgAsnGluGlyGlySer 539
DB 551 TGGGACACGAGAACTGAGGTTGCAGATAAGTCATGTTTACACAGGAATGAAGCTGGGTCA 610
QY 540 LysTrpGlyTrpTrpArgArgValTrpTrpThrLeuLysProTrpLysAlaAsnTrpThr 559
DB 611 AAGTACGGGTAATGCTGTCAGAGTGGATGACACACATTCCTCTGCAAGCAATGATACC 670
QY 560 MetTrpGlyLysLeuSerTrpGlnGlyLysSerTrpAsnLeuProTrpLysGlyArgile 579
DB 671 ATGTGTGGGAAGTGTGTTCTGTCAAGTGGTGGGATAATTTGCCCTGGAAAGACGGATA 730
QY 580 ValThrTrpLeuThrTrpLysThrTrpTrpProGluTrpThrSerGlnGluLeuGlyMet 599
DB 731 GTGACTTTTCTGACATGTAAACCATTTGATCTCTGAAGACACAAAGTCAAGAAATAGGCATG 750

QY 600 VallalalenGlyThrLysTrpGlyTrpAsnLysValTrpIleAsnAlaGluTrpValTrp 619
 Db 791 GTGCCAATGGAACCTAAGTGTGGGATACAAAGGTTTCATTAATGCAAGATGTGGAT 850
 QY 620 IleGluLysAlaLysTrpSerThrAsnTrpSerSerLysTrpLysGlyHisAlaValTrp 639
 Db 851 ATTGAGAAAGCTACAAATCAACCAATGCTCACTAAGTGCMAAGGACATGCTGTGTGT 910
 QY 640 TrpHisGluLeuGlnTrpGlnTrpGluGluGlyTrpIleProProTrpTrpTrpTrpSer 659
 Db 911 GACCATGAGCTCCAGTGTCAATGTGAGGAAGATGATCCCTCCCGACTGCGATGACTCC 970
 QY 660 SerValValTrpHisTrpSerIleValValGlyValLeuTrpProMetAlaValIleTrp 679
 Db 971 TCAGTGGCTCTCCACTCTCCATTTGTTGGTGGGGTGTGTTCCCAATGGCGTCAATTTT 1030
 QY 680 ValValAlaAlaMetValIleArgHisGlnSerSerArgGluLysGlnLysLysTrpGln 699
 Db 1031 GTGTGGTGTCTATCGGTAATCGGCACCAGAGCTCCAGAGAAAGSCAGAAAGATCAG 1090
 QY 700 ArgProLeuSerThrThrGlyThrArgProHisLysGlnLysArgLysProGlnMetVal 719
 Db 1091 AGGCCACTATCTCCACTGGCCAGCCAGGCCACACAAACAGAGAGAAACCCAGATGTA 1150
 QY 720 LysAlaValGlnProGlnGluMetSerGlnMetLysProHisValTrpLeuProVal 739
 Db 1151 ARGCTGTTCACCCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1210
 QY 740 GluGlyAsnGluProProAlaSerTrpHisLysTrpTrpAsnAlaLeu2ProProThrVal 759
 Db 1211 GAAGCAATGAGCCCGCCAGCCCTTTTCATAAAGACACAAAGCGCACTTCCCTACTGT 1270
 QY 760 TrpLysTrpAsnProMetSerThrProLysTrpSerAsnProLysAla 775
 Db 1271 TTCAGGATAATCCAAATGCTACACCTAAGGACTCAATCCAAAAGCA 1318
 RESULT 11
 ABX90528
 ID ABX90528 standard; cDNA; 2049 BP.
 XX
 AC ABX90528;
 XX
 DT 02-MAY-2003 (first entry)
 XX
 DE Human cDNA from novel ADAM gene 3.
 XX
 KW Human; ss; Gene; ADAM; a dystrophin and metalloprotease; cancer; stroke;
 KW neural disease; cardiovascular disease; heart disease; atherosclerosis;
 KW angina; reproductive disease; gastrointestinal disease; Crohn's disease;
 KW endocrine disease; renal disease; central nervous system disease; AIDS;
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease; asthma;
 KW schizophrenia; inflammatory disease; immunological disease; osteoporosis;
 KW rheumatoid arthritis; inflammatory bowel disease; bone formation disease;
 KW musculo-skeletal disease; angiogenic disease; blood related disorder;
 KW respiratory neoplasia; thrombosis; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN US2002173640-A1.
 XX
 PD 21-NOV-2002.
 XX
 PF 19-APR-2002; 2002US-00125452.
 XX
 PR 27-MAY-1999; 99US-0136388P.
 PR 09-JUL-1999; 99US-0142930P.
 PR 28-JAN-2000; 2000US-0178717P.
 PR 25-MAY-2000; 2000WO-US14308.
 PR 21-SEP-2000; 2000US-0234222P.
 PR 16-NOV-2000; 2000US-00712907.
 PR 19-SEP-2001; 2001US-00395504.
 XX

PA (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Ni J, Hastings GA, Shi Y, Wei P;

XX P-PSDB; ABU60627.

DR New isolated ADAM (a disintegrin and metalloprotease domain) polypeptide
 XX and encoding polynucleotide useful for diagnosing and treating ADAM-
 PT related pathologies e.g. cancer, blood disorders, inflammatory and immune
 PT disorders and thrombosis.

XX Claim 1; Page 121; 147pp; English.

XX The invention relates to a new isolated polypeptide, comprising a
 CC sequence at least 95% identical to any one of 7 fully defined sequences
 CC given in the specification, or its fragment, epitope or variant. The
 CC polypeptides are ADAM proteins (a dystrophin and metalloprotease). Also
 CC included are the nucleic acids encoding the ADAM proteins, a vector
 CC comprising the isolated nucleic acid, a host cell comprising the vector,
 CC a recombinant host cell comprising the nucleic acid operably linked to a
 CC heterologous regulating element which controls gene expression, an
 CC isolated antibody that specifically binds to the ADAM protein,
 CC identifying a binding partner to ADAM (comprising contacting ADAM with a
 CC binding partner, and determining whether the binding partner affects at
 CC activity of the polypeptide) and screening for molecules which modify
 CC activities of ADAM (comprising contacting the polypeptide with a compound
 CC suspected of having agonist or antagonist activity, and assaying for
 CC activity of the polypeptide). The methods and compositions of the present
 CC invention are useful for diagnosing, preventing, ameliorating and/or
 CC treating diseases associated with the aberrant expression or activity of
 CC the ADAM polypeptide, such as neural, cardiovascular (e.g. heart disease,
 CC atherosclerosis, stroke, angina), reproductive, gastrointestinal (e.g.
 CC Crohn's disease), endocrine, renal, central nervous system (e.g.
 CC Alzheimer's, Parkinson's, Huntington's, schizophrenia), inflammatory,
 CC immunological (e.g. AIDS, rheumatoid arthritis, asthma, inflammatory
 CC bowel disease) bone formation, musculo-skeletal (e.g. osteoporosis),
 CC angiogenic and blood related disorders, and respiratory neoplasias,
 CC cancer and thrombosis. Many more diseases and disorders are listed in the
 CC specification. They can also be used in chromosome identification,
 CC differential identification of tissue and cell types, screening assays
 CC and molecular weight markers. The present sequence encodes an ADAM
 CC protein

SQ Sequence 2049 BP; 671 A; 368 C; 445 G; 565 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.14e-81 Length: 2049
 Score: 1576.00 Matches: 341
 Percent Similarity: 82.84% Conservative: 21
 Best Local Similarity: 78.03% Mismatches: 74
 Query Match: 33.45% Indels: 1
 Db: 7 Gaps: 0

US-10-726-148A-15 (1-775) x ABX90528 (1-2049)

QY 340 GluMetGlyHisAsnTrpGlyMetTrpHisTrpTrpTrpTrpTrpTrpTrpSerThr 359
 Db 11 GAATGGGCCAACAACTTGGGATGTTCTGACGACTATCTTGCAGATGTCCTCTACA 70
 QY 360 IleTrpValMetTrpLysAlaLeuSerTrpTrpTrpTrpTrpTrpTrpTrpTrpSerThr 379
 Db 71 ATATGTGTGATGGACAAAGCACTGAGCTTCTATATACCACAGACTTCAGTTCCTGCAGC 130
 QY 380 ArgLeuSerTrpTrpLysTrpGluTrpLysLeuSerAsnTrpLeuTrpAsnAlaPro 399
 Db 131 CGTCCAGCTATGACAAAGTATTTTGAAGATAAATATCAAAATTCGCTCTTAAATGCTCCA 190
 QY 400 LeuProThrTrpIleIleSerThrProIleTrpGlyAsnGlnLeuValGluMetGlyGlu 419
 Db 191 TTGCTACAGATATCATATCCATCCATCCATTCGAGACCAAGTGTGGTGAATGGAGAG 250
 QY 420 TrpTrpTrpTrpGlyThrSerGluGluTrpThrAsnIleTrpTrpTrpAlaLysThrTrp 439

```
Db 251 GACTGTGATTTGGGACATCTGAGSAATGTACCAATATTTCTGTGATGCTTAAGACATGT 310
Qy 440 LysIleLysAlaThrTrpGlnTrpAlaLeuGlyGluTrpTrpGluLysTrpGlnTrpLys 459
Db 311 AAAATCAAGCAACTTTTCAATGTGCAATAGAGAAATCTTTGTAATAATGCAATTTAAA 370
Qy 460 LysAlaGlyMetValTrpArgProAlaLysTrpGluTrpTrpLeuProGluMetTrpAsn 479
Db 371 AAGGCTGGATGGTGTGCAGACACCAAAAGATGAGTCGCACCTGCCCTGAAATGTTAAT 430
Qy 480 GlyLysSerGlyAsnTrpProTrpTrpArgTrpGlnValAsnGlyTrpProTrpHisHis 499
Db 431 GGTAATCTGTAATGTCTCGATGATAGATTCGAAGTCAATGGCTTCCCTTGCCATCAC 490
Qy 500 GlyLysGlyHisTrpLeuMetGlyThrTrpProThrLeuGlnGluTrpThrGluLeu 519
Db 491 GGGGAAGGCCCATCTGTTGATGGGACATGCCCCCACTGCGAGGACAGTGCACAGAGCTG 550
Qy 520 TrpGlyProGlyThrGluValAlaTrpLysSerTrpTyrAsnArgAsnGluGlyGlySer 539
Db 551 TGGGACCAAGCACTGAGTTGCGATGAAGTCATGTTACACAGGAATGAAGGTGGGTCA 610
Qy 540 LysTyrGlyTyrTrpArgArgValTrpTrpThrLeuIleProTrpLysAlaAsnTrpThr 559
Db 611 AAGTACGGGTACTGTGCGAGAGTGGATGCACACATCTATCCCTGCAAGCAATGATACC 670
Qy 560 MetTrpGlyLysLeuTrpTrpGlnGlyGlySerTrpAsnLeuProTrpLysGlyArgIle 579
Db 671 ATGTGTGGGAAGTTGTTCTGTCAAGGTGGTGGATTAATTTGCCCTGGAAGGACGGATA 730
Qy 580 ValThrTrpLeuThrTrpLysThrTrpTrpProGluTrpThrSerGlnGluIleGlyMet 599
Db 731 GTGACTTTCGACATGTAAACATTTGCTGTAAGACACAAAGTCAAGAAATAGGCATG 790
Qy 600 ValAlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpIleAsnAlaGluTrpValTrp 619
Db 791 GTGGCCAATGGAATAGTGTGGGATAAACAAGTTTGCATTATGTCAGAGATGTGTGAT 850
Qy 620 IleGluLysAlaThrLysSerTrpAsnTrpSerSerLysTrpLysGlyHisAlaValTrp 639
Db 851 ATTGAGAAAGCCTCAAAATCAACCAATGTCTCATCTAAGTCCAAAGGACATGCTGTGCT 910
Qy 640 TrpHisGluLeuGlnTrpGlnTrpGluGlyTrpIleProTrpTrpTrpTrpSer 659
Db 911 GACCATGAGCTCCAGTGTCAATGTGAGAAAGATGGATCCCTCCGACTGCGATGACTCC 970
Qy 660 SerValValTrpHisTrpSerIleValValGlyValLeuTrpProMetAlaValIleTrp 679
Db 971 TCAGTGGTCTTCCACTTCTCCATTTGCTGTGGTGGTGTGTTCCCAATGGCGTCAATTTT 1030
Qy 680 ValValValAlaMetValIleArgHisGlnSerSerArgGluLysGlnLysLysTrpGln 699
Db 1031 GTGGTGGTGTGATGGTAATCCGGCCACAGAGCTCCAGAGAAAGGACGAGAAAGATCAG 1090
Qy 700 ArgProLeuSerThrThrGlyThrArgProHisLysGlnLysAlaGlyProGlnMetVal 719
Db 1091 AGGCCACTATCTACCATCTGGCCACAGGCCACACAAACAGAGAGGAACCCACAGTGTGA 1150
Qy 720 LysAlaValGlnProGlnGluMetSerGlnMetLysProHisValTyr-TrpLeuProVa 739
Db 1151 AAGGCTGTTCACCCCAAGAGATGAGTCAGATGAAGCCCCCATGTGTATGATCTCTGCCAGT 1210
Qy 739 lGluGlyAsnGluProProAlaSerTrpHisLysTrpThrAsnAlaLeuProProThrVa 759
Db 1211 AGAAGGCAATGAGCCCCCAGCCTCTTTTCATAAAGACACACATGTGTATGATGACCTACTGT 1270
Qy 759 lTrpLysTrpAsnProMetSerThrProLysTrpSerAsnProLysAla 775
Db 1271 TTTCAAGGATATCAATGTCTACACCTAAGGACTCAATATCCAAAGCA 1319
```

RESULT 12
AAC98231

```
ID AAC98231 standard; cDNA; 1652 BP.
XX AAC98231;
AC 09-MAR-2001 (first entry)
DT Human colon cancer antigen nucleotide sequence SEQ ID NO:241.
XX
DE Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnerary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder; ss.
XX
OS Homo sapiens.
XX WO200055351-A1.
PN 21-SEP-2000.
PD 08-MAR-2000; 2000WO-US005883.
PF 12-MAR-1999; 99US-0124270P.
PR (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Ruben SM;
XX WPI; 2000-587534/55.
XX P-PSDB; AAB53474.
DR Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer.
XX
PS Claim 1; Page 660-661; 2104pp; English.
XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnerary, nephrotropic, antiinfective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins may
CC also be used to prevent diseases such as neural disorders, immune system
CC disorders, muscular disorders, reproductive disorders, gastrointestinal
CC disorders, wounds, renal disorders, infectious diseases, and
CC cardiovascular disorders. AAC98764 to AAC98772 and AAB54007 represent
CC sequences used in the exemplification of the present invention
XX
SQ Sequence 1652 BP; 560 A; 289 C; 333 G; 464 T; 0 U; 6 Other;
Alignment Scores:
Pred. No.: 2.75e-40 Length: 1652
Score: 854.00 Matches: 179
Percent Similarity: 86.64% Conservative: 9
Best Local Similarity: 82.49% Mismatches: 29
Query Match: 18.12% Indels: 0
DB: 3 Gaps: 0
US-10-726-148A-15 (1-775) x AAC98231 (1-1652)
Qy 559 ThrMetTrpGlyLysLeuTrpTrpGlnGlyGlySerTrpAsnLeuProTrpLysGlyArg 578
Db 251 ACCATGTGTGGAAAGTTGTTCTGTCAAGTGGTGGGATTAATTTGCCCTGGAAGACAG 310
Qy 579 IleValThrTrpLeuThrTrpLysThrTrpProGluTrpThrSerGlnGluIleGly 598
```

Db 311 ATAGTGACCTTCTCGACATGTAACACATTTGATCCTCGAGACACACAACTCAGAAATARGC 370
QY 599 MetValAlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpLeuAlaGluTrpVal 618
Db 371 ATGATGCGCAATGAACTAAGTGTGGCGATAACAAGGTTTGGCATKATGAGAAATGTGTG 430
QY 619 TrpIleGluLysAlaValLysSerThrAsnTrpSerSerLysTrpLysGlyHisAlaVal 638
Db 431 GATATTGAGAAAGCCTCAAAATCAACCAATGCTCATCAAGTCAAAAGGACATGCTGTG 490
QY 639 TrpTrpHisGluLeuGlnTrpGlnTrpGlnLysGluGlyTrpIleProTrpTrpTrpTrp 658
Db 491 TGTGACCATGAGCTCCAGTCCATGTCGAGGAGGATGGATCCCTCCGACTCGCATGAY 550
QY 659 SerSerValValTrpHisTrpSerIleValValGlyValLeuTrpProMetAlaValIle 678
Db 551 TCTCAGTGTCTTCCACTTCTCCATGTGGTGTGGGTGCTGTCTCCCAATGGCGGTCAAT 630
QY 679 TrpValValValAlaMetValIleArgHisGlnSerSerArgGluLysGlnLysLysTrp 698
Db 611 TTGTGTGTGTGTCTATGTTAATCCGTCACAGAGCTCCAGAGAAAGCAGAGAAAGAT 670
QY 699 GluArgProLeuSerThrThrGlyThrArgProHisLysGlnLysArgLysProGlnMet 718
Db 671 CAGAGGCCATATCTACCACTGGCACCAGGCCACACAAACAGAGAGGAAACCCCGATG 730
QY 719 ValLysAlaValGlnProGlnGluMetSerGlnMetLysProHisValTrpTrpLeuPro 738
Db 731 GTAAAGGCTGTTCAACCCCAAGAGATGAGTCAGATGAAGGCCCATGTGTATGATCTGCCA 790
QY 739 ValGluGlyAsnGluProProAlaSerTrpHisLysTrpThrAsnAlaLeuProProThr 758
Db 791 GTAGAAGCAATGAGCCCCAGCCCTCTTTTCATAAAGACACAAACGACACTTCCCTACT 850
QY 759 ValTrpLysTrpAsnProMetSerThrProLysTrpSerAsnProLysAla 775
Db 851 GTTTTCAAGGATAATCAATGTCTACACCTAAGGACTCAAAATCCAAAAGCA 901

RESULT 13

AAAC91180
ID AAC9118C standard; DNA; 1704 BP.
XX AC AAC9118C;
XX AC AAC9118C;
XX 20-MAR-2001 (first entry)
XX DE ADAM gene #4.
XX ADAM; disintegrin; metalloprotease; food additive; breast cancer;
XX ovarian; ss.
XX Homo sapiens.
XX WO2000073323-A2.
XX 07-DEC-2000.
XX 25-MAY-2000; 2000WO-US014308.
XX 27-MAY-1999; 99US-0136388P.
XX 09-JUL-1999; 99US-0142930P.
XX 28-JAN-2000; 2000US-0178717P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Wei P, Ni J, Hastings GA, Shi Y;
XX WPI; 2001-016507/02.
XX Seven nucleic acid molecules encoding ADAM polypeptides containing a
XX disintegrin and metalloprotease domain, useful in the prevention,
XX treatment and diagnosis of cancer, immune disorders, cardiovascular
XX disorders and neurological diseases.

XX PS Claim 1; Page 267-268; 287pp; English.
XX CC The present invention relates to seven members of the ADAMs (proteins
XX which contain A Disintegrin And Metalloprotease domain) protein family.
XX CC The ADAMs proteins and DNA may be used to treat disease, as a food
XX CC additive or preservative, for chromosome identification, as probes for
XX CC diagnosing a disorder related to the female reproductive system,
XX CC particularly breast and/or ovary cancer. They are also useful in the gene
XX CC therapy of breast and ovarian cancer
XX SQ Sequence 1704 BP; 586 A; 293 C; 332 G; 493 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2,55e-38 Length: 1704
Score: 820.00 Matches: 174
Percent Similarity: 84.33% Conserv: 9
Best Local Similarity: 80.18% Mismatches: 26
Query Match: 17.40% Indels: 8
DB: 4 Gaps: 1

US-10-726-148A-15 (1-775) x AAC91180 (1-1704)

QY 559 ThrMetTrpGlyLysLeuTrpTrpGlnGlySerTrpAsnLeuProTrpLysGlyArg 578
Db 259 ACCATGTGTGGGAAGTTTCTGTCTCAAGTGGGTCCGATAATTTGCCCTCGAAAGGACGG 318
QY 579 IleValThrTrpLeuThrTrpLysThrTrpTrpProGluTrpThrSerGlnGluIleGly 598
Db 319 ATAGTGACTTTTCTCGACATGTAAACACATTTGATCCTCGAAGACACAACTCAAGAAATAGGC 378
QY 599 MetValAlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpIleAsnAlaGluTrpVal 618
Db 379 ATGATGCGCAATGGAACATAAGTGTGGCGATACAAAGTTTGCATTAATGCAGATGTGTG 438
QY 619 TrpIleGluLysAlaValLysSerThrAsnTrpSerSerLysTrpLysGlyHisAlaVal 638
Db 439 GATATTGAGAAAGCCTCAAAATCAACCAATTCCTCATCTAAGTCAAAAGGACATGCTGTG 498
QY 639 TrpTrpHisGluLeuGlnTrpGlnTrpGluGluGlyTrpIleProTrpTrpTrpTrp 658
Db 499 TGTGACCATGAGCTCCAGTGTCAATGTGAGGAGGATGGATCCCTCCGACTCGCATGAC 558
QY 659 SerSerValValTrpHisTrpSerIleValValGlyValLeuTrpProMetAlaValIle 678
Db 559 TCTCAGTGTCTTCCACTTCTCCATGTGGTGGGTGCTGTCTCCCAATGGCGGTCAAT 618
QY 679 TrpValValValAlaMetValIleArgHisGlnSerSerArgGluLysGlnLysLysTrp 698
Db 619 TTGTGTGTGTGTCTATGTTAATCCGTCACAGAGCTCCAGAGAAAGCAGAGAAAGTT 678
QY 699 GluArgProLeuSerThrThrGlyThrArgProHisLysGlnLysArgLysProGlnMet 718
Db 679 CAGAGGCCACTATCTACCACTGGCACCCAGGCCACACAAACAGAGAGGAAACCCCGATG 738
QY 719 ValLysAlaValGlnProGlnGluMetSerGlnMetLysProHisValTrpTrpLeuPro 738
Db 739 -----ATGAGTCAGATGAAGCCCCCATGTGTATGATCTGCCA 774
QY 739 ValGluGlyAsnGluProProAlaSerTrpHisLysTrpThrAsnAlaLeuProProThr 758
Db 775 GTAGAAGCAATGAGCCCCCAGCGCTCTTTTCATAAAGACACAAACGACACTTCCCTACT 834
QY 759 ValTrpLysTrpAsnProMetSerThrProLysTrpSerAsnProLysAla 775
Db 835 GTTTTCAAGGATAATCAATGTCTACACCTAAGGACTCAAAATCCAAAAGCA 885

RESULT 14

ABX90529
ID ABX90529 standard; CDNA; 1704 BP.
XX AC ABX90529;
XX AC ABX90529;

DT 02-MAY-2003 (first entry)

DE Human CDNA from novel ADAM gene 4.

XX Human; ss; gene; ADAM; a dystrophin and metalloprotease; cancer; stroke;
 KW neural disease; cardiovascular disease; heart disease; atherosclerosis;
 KW angina; reproductive disease; gastrointestinal disease; Crohn's disease;
 KW endocrine disease; renal disease; central nervous system disease; AIDS;
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease; asthma;
 KW schizophrenia; inflammatory disease; immunological disease; osteoporosis;
 KW rheumatoid arthritis; inflammatory bowel disease; bone formation disease;
 KW musculo-skeletal disease; angiogenic disease; blood related disorder;
 KW respiratory neoplasia; thrombosis; gene therapy.

XX Homo sapiens.

XX US2002173640-A1.

XX 21-NOV-2002.

XX 19-APR-2002; 2002US-00125452.

XX 27-MAY-1999; 99US-0136388P.

XX 03-JUL-1999; 99US-0142930P.

XX 28-JAN-2000; 2000US-01787-7P.

XX 25-MAY-2000; 2000WO-US214308.

XX 21-SEP-2000; 2000US-0234222P.

XX 18-NOV-2000; 2000US-00712907.

XX 19-SEP-2001; 2001US-00955504.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Ni J, Hastings GA, Shi Y, Wei P;

XX WPT; 2003-255274/25.

XX P-PSDB; ABU60628.

XX New isolated ADAM (a disintegrin and metalloprotease domain) polypeptide
 PT and encoding polynucleotide useful for diagnosing and treating ADAM-
 PT related pathologies e.g. cancer, blood disorders, inflammatory and immune
 PT disorders and thrombosis.

XX Claim 1; Page 122; 147pp; English.

XX The invention relates to a new isolated polypeptide, comprising a
 CC sequence at least 95% identical to any one of 7 fully defined sequences
 CC given in the specification, or its fragment, epitope or variant. The
 CC polypeptides are ADAM proteins (a dystrophin and metalloprotease). Also
 CC included are the nucleic acids encoding the ADAM proteins, a vector
 CC comprising the isolated nucleic acid, a host cell comprising the vector,
 CC a recombinant host cell comprising the nucleic acid operably linked to a
 CC heterologous regulating element which controls gene expression, an
 CC isolated antibody that specifically binds to the ADAM protein,
 CC identifying a binding partner to ADAM (comprising contacting ADAM with a
 CC binding partner, and determining whether the binding partner affects an
 CC activity of the polypeptide) and screening for molecules which modify
 CC activities of ADAM (comprising contacting the polypeptide with a compound
 CC suspected of having agonist or antagonist activity, and assaying for
 CC activity of the polypeptide). The methods and compositions of the present
 CC invention are useful for diagnosing, preventing, ameliorating and/or
 CC treating diseases associated with the aberrant expression or activity of
 CC the ADAM polypeptide, such as neural, cardiovascular (e.g. heart disease,
 CC atherosclerosis, stroke, angina), reproductive, gastrointestinal (e.g.
 CC Crohn's disease), endocrine, renal, central nervous system (e.g.
 CC Alzheimer's, Parkinson's, Huntington's, schizophrenia), inflammatory,
 CC immunological (e.g. AIDS, rheumatoid arthritis, asthma, inflammatory
 CC bowel disease) bone formation, musculo-skeletal (e.g. osteoporosis),
 CC angiogenic and blood related disorders, and respiratory neoplasias,
 CC cancer and thrombosis. Many more diseases and disorders are listed in the
 CC specification. They can also be used in chromosome identification,
 CC differential identification of tissue and cell types, screening assays
 CC and molecular weight markers. The present sequence encodes an ADAM
 CC protein

XX Sequence 1704 BP; 586 A; 293 C; 332 G; 493 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2.55e-38 Length: 1704
 Score: 820.00 Matches: 174
 Percent Similarity: 84.33% Conservative: 9
 Best Local Similarity: 80.18% Mismatches: 26
 Query Match: 17.40% Indels: 8
 DB: 7 Gaps: 1

US-10-726-148A-15 (1-775) x ABX90529 (1-1704)

QY 559 ThrMetTrpGlyLysLeuTrpTrpGlnGlySerTrpAsnLeuProTrpLysGlyArg 578
 DB 259 ACCATGTGTGGAAAGTTGTTCTGTCAAGTGGTGGGATTAATTTGCCCTGGAAAGACGS 318
 QY 579 IleValThrTrpLeuThrTrpLysThrTrpTrpProGluTrpThrSerGlnGluLeGly 598
 DB 319 ATAGTGACTTCTGACATGTAAACATTTGATCTCTGAAGACACAAAGTCAAGAAATAGGC 378
 QY 599 MetValAlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpLleAsnAlaGluTrpVal 618
 DB 379 ATGTGGCCAAATGGAACCTAAGTGTGGCGATAACAAAGGTTTGCAATTAATGCAAGATGTG 438
 QY 619 TrpLleGluLysAlaLysLysSerThrAsnTrpSerSerLysTrpLysGlyHisAlaVal 638
 DB 439 GATATTGAGAAAGCCTCAAAATCAACCAATTCCTCTCTAGTGCAGAAAGCATCTCTGTG 498
 QY 639 TrpTrpHisGluLeuGlnTrpGlnTrpGluGlyTrpLleProProTrpTrpTrp 658
 DB 499 TGTGACCATGAGTCTCCAGTGTCAATGTGAGGAAGATGGATCCCTCCCGACTGCCGATGAC 558
 QY 659 SerSerValValTrpHisTrpSerLleValValGlyValLeuTrpProMetAlaValIle 678
 DB 559 TCCTCAGTGGTCTTCCACTTCTCCATTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 618
 QY 679 TrpValValValAlaMetValLleArgHisGlnSerSerArgGluLysGlnLysLysTrp 698
 DB 619 TTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 678
 QY 699 GlnArgProLeuSerThrThrArgProHisLysGlnLysGlnLysArgLysProGlnMet 718
 DB 679 CAGAGGCCACTATCTACCTGCGACCCAGCCACACAAACAGAGAGAGAAACCCAGATG 738
 QY 719 ValLysAlaValGlnProGlnGlnMetSerGlnMetLysProHisValTrpLleuPro 738
 DB 739 -----ATGAGTCAGATGAAGCCCATGTGTGTATGATCTGCCA 774
 QY 739 ValGluGlyAsnGluProProAlaSerTrpHisLysTrpThrAsnAlaLeuProProThr 758
 DB 775 GTAGAGGCATGAGAGCCCGCCAGCCCTCTTTTATTAAGACACAAAGCGACTTCCCCTACT 834
 QY 759 ValTrpLysTrpAsnProMetSerThrProLysTrpSerAsnProLysAla 775
 DB 835 GTTTTCAAGGATAATCCAATGCTACACCTAAGGACTCAAATCCAAAGCA 885

RESULT 15

ABK10644

ID ABK10644 standard; cDNA; 2334 BP.

AC ABK10644;

DT 05-JUN-2002 (first entry)

XX Pigmy rattlesnake disintegrin protease (DP) zsnk10 gene sequence #2.
 DE disintegrin homologue; ZSNK10; ZSNK11; ZSNK12; platelet aggregation;
 KW disintegrin homologue; ZSNK10; ZSNK11; ZSNK12; platelet aggregation;
 KW apoptosis; neurogenesis; myogenesis; immunologic recognition;
 KW tumour formation; brain disorder; heart infarction; Alzheimer's disease;
 KW multiple sclerosis; congestive heart failure; ischaemic reperfusion;
 KW coagulation disorder; degenerative disease.

OS Sistrurus miliarius.
 FH Key Location/Qualifiers
 FT CDS 95..1930
 FT FT /tag= a
 FT FT /product= "Disintegrin protease zsnk10 protein #2"
 FT FT /tag= b
 FT FT /tag= d
 FT FT /tag= c
 FT FT /note= "Encodes pro-peptide domain of zsnk10 protein #2"
 FN #C200212283-A2.
 XX
 PD 14-FEB-2002.
 XX
 PF 02-AUG-2001; 2001-NO-US024536.
 XX
 PR 03-AUG-2000; 2000US-0222564P.
 XX
 PA (ZYMO) ZYMOGENETICS INC.
 XX
 PI Fox BA, Sheppard PO;
 XX
 DR WPI; 2002-241734/29.
 DR P-ESDB; AAU76794.
 XX
 XX New polypeptides comprising disintegrin homologs, useful modulating cell-cell interactions, particularly for treating or preventing e.g. Alzheimer's disease, tumor, multiple sclerosis, heart failure or ischemic reperfusion.
 PS Disclosure; Page 91-95; 108pp; English.
 XX
 CC The present invention relates to new disintegrin homologues, defined as ZSNK10, ZSNK11, and ZSNK12, which are members of Disintegrin proteases. The disintegrin polypeptides of the invention are useful for modulating cell-cell interactions e.g. platelet aggregation, apoptosis, neurogenesis, myogenesis, immunologic recognition or tumour formation. In particular, the polypeptides are useful for treating or preventing disorders associated with brain or heart infarction e.g. Alzheimer's disease, tumour formation, multiple sclerosis, congestive heart failure, ischemic reperfusion, coagulation disorders or degenerative diseases. The present nucleic acid sequence encodes the disintegrin protease (DP) homologue zsnk10 protein #2 of the invention. Zsnk10 is one of several DP homologues of the invention
 SQ Sequence 2334 BP; 735 A; 492 C; 506 G; 600 T; 0 U; 1 Other;
 Alignment Scores:
 Pred. No.: 6.67e-35 Length: 2334
 Score: 763.00 Matches: 226
 Percent Similarity: 51.39% Conservative: 106
 Best Local Similarity: 34.98% Mismatches: 280
 Query Match: 16.19% Indels: 34
 DB: 6 Gaps: 13
 US-10-726-148A-15 (1-775) x ABK10644 (1-2334)
 QY 1 MetLeuGlnGlyLeuLeuProValSerLeuLeuSerVal-----AlaVal 16
 Db 95 ATGATCCAGATTCTCTTG---GTGACTATATGCTTAGCAGGTTCTCTTATCAAGGGAGC 151
 QY 17 SerAlaIleLysGluLeuProGlyValLysLysTyrGluValValTyrProIleArgLeu 36
 Db 152 TCTATATCTCGGAATCTGGGAACGTGAATGATTTATGAAGTAGTGTATACACGAAGATC 211
 QY 37 HisProLeuHisLysArgGluAlaLysGluProGluGlnGlnGluLysTyrGluThrGlu 56
 Db 212 ACTGCATTGCCCCAAGAGGAGAGCT-----CAGCCAAAGATGAAGAGGCC 256

QY 57 LeuLysTyrLysMetThrIleAsnGlyLysIleAlaValLeuTyrLeuLysLysLysLys 76
 Db 257 ATGCAATATGAATTAAAGATCAACGGAGACCACTGGTCTCTTCACCTCGAATAAATAA 316
 QY 77 AsnLeuLeuAlaProGlyTyrThrGluTyrTyrAsnSerThrGlyLysGluIleThr 96
 Db 317 AGACTTTTTCAAAAGATTACAGCAGACTCATTTATCCCTGATGGCAGACAATATACA 376
 QY 97 ThrSerProGlnIleMetTyrTyrTyrGlnGlnHisIleLeuAsnGluLysVal 116
 Db 377 ACATACCCCATGATTGAGGATCACTGCTATTATCATGAGCATCCAGAAATGATGCTGAC 436
 QY 117 SerTrpAlaSerIleSerThrTrpArgGlyLeuArgGlyTyrTrpSerGlnGlyTrpGln 136
 Db 437 TCAACTGCAAGCATCAGTGCATCAACGGTTTGAAGGACATTTCAAGCTTCAAGGGGAG 496
 QY 137 ArgTyrTrpIleGluProLeuSerProIleHisArgTrpGlyGlnGluHisAlaLeuTrp 156
 Db 497 ATGTACCTCATTTGAACCTTGAAGCTTCCGAC-----AGTGAAGCCCATGAGTCTAC 550
 QY 157 LysTyrAsnProTrpGluLysAsnTyrTrpSerThrTrpGlyMetTrpGlyValLeuTrp 176
 Db 551 AAATATGAAAAACATAGAAAAGAGGATGAGCCGCC---AAAATGTGTGGGTAAACC--- 604
 QY 177 AlaHisTrpLeuGlnGlnAsnIleAlaLeuProAlaThrLysLeuValLysLeuLysTrp 196
 Db 605 CAGAAITGGGAATCATATGAG-----CCCATCAAAAGGCCCTTTCAGTTAAATCTT 655
 QY 197 ArgLysValGln-----GluHisGluLysTyrIleGluTyrTyrLeuValLeu 212
 Db 656 ACTCTGACACACACAGCATACTTGGATGCCAAAATAACGTGAGTTGTGTAGTCTTG 715
 QY 213 TrpAsnGlyGluTrpLysArgTyrAsnGluAsnGlnTrpGluIleArgLysArgValTrp 232
 Db 716 GACCATGGAATGTACACAAAATACAAGATGATTTAGATAAGATAAACAAGATATAT 775
 QY 233 GluMetAlaAsnTyrValAsnMetLeuTyrLysLysLysLeuAsnThrHisValAlaLeuVal 252
 Db 776 GAATGTCACACTGAATGAGATTTACATACCTTTGAATATTCGTGTGCGACTGTT 835
 QY 253 GlyMetGluIleTrpThrLysLysLysLysLysIleThrProAsnAlaSerThr 272
 Db 836 CACCTAGAAATTTGGTCCACACAGAGATTGATTAATGTGTCATCAGCAGCAGGTGATACT 895
 QY 273 LeuGluAsnTrpSerLysTrpArgGlySerValLeuSerArgArgLysArgHisTrpIle 292
 Db 896 TTGGGCTCATTTGGAGAATGGAGAGACAGATTTGCTGAGGCCACAAAAGTCATGATAAT 955
 QY 293 AlaGlnLeuIleThrAlaThrGluLeuAlaGlyThrThrValGlyLeuAlaTrpMetSer 312
 Db 956 GCTCAGTTACTCAGCACCCTGACTTCGATGGAGACACTGTAGGATTGGCTTATATAGC 1015
 QY 313 ThrMetTrpSerPro---TyrSerValGlyValValGlnTrpHisSerTrpAsnLeuLeu 331
 Db 1016 AGCATGCGCCACCGAGCAGTCTGTAGGAGTTATTTCAGCAACATAGCACACAATCTC 1075
 QY 332 ArgValAlaGlyThrMetAlaHisGluMetGlyHisAsnTrpGlyMetTrpHisTrpTrp 351
 Db 1076 TTGATGGCAGTTACAATGGCCCATGAGATGGGTCATAATCTGGGCATGAGTCATGATGGA 1135
 QY 352 TyrSerTrpLysTrpProSerThrIleTrpValMetTrpLysAlaLeuSerTrpTyrIle 371
 Db 1136 AATCAGTGTCTATGTTGGTGTCTCTCTCGTCATTTATGCTGAAGACTAAGCCCAACCT 1195
 QY 372 ProThrTrpTrpSerSerTrpSerArgLeuSerTyrTrpLysTrpTrpGluTrpLysLeu 391
 Db 1196 TCCACACAGTTTCAGGATGCTAGTGAGGAATATTGTGCGAGCTATCTTAAATACTGAGA 1255
 QY 392 SerAsnTrpLeuTrpAsnAlaProLeuProThrTrpIleIleSerThrProIleTrpGly 411
 Db 1256 CCACAATGCATTCTCAATGAACCTTGTCTGACAGATATTGTTCACTCCAGTTTGTGGA 1315

```
Qy 412 AsnGlnLeuValGluMetGlyGluTrpTrpTrpGlyThrSerGluGluTrpThrAsn 431
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1316 AATGAACITTTGGAGGAGGAGAGAGAGTGAAGTGAAGTCTCTCTCGAAACCTGTGAGAAT 1375
Qy 432 IleTrpTrpTrpAlaLysThrTrpLysIleLysAlaThrTrpGlnTrpAlaLeuGlyGlu 451
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1376 CCATGCTGTGATGCTGCAACGCTAAACAGCACCCAGGGTCACAGTGTGCAAAAGGAGCTG 1435
Qy 452 TrpTrpGluLysTrpGlnTrpLysLysAlaGlyMetValTrpArgProAlaLysTrpGlu 471
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1436 TGTGTGACCACTGACAGATTTAAAGGGGCGCAGAACAGAAATGCCGGCGCAGCAAGGATGAC 1495
Qy 472 TrpTrpLeuProGluMetTrpAsnGlyLysSerGlyAsnTrpTrpTrpTrpArgTrpGln 491
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1496 TGTGACATGGCTGATCTCTGCACTGGCCAACTGCTAAGTGTCCACGGATCGTCCCAA 1555
Qy 492 ValAsnGlyTrpProTrpHisGlyLysGlyHisTrpLeuMetGlyThrTrpProThr 511
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1556 AGGAATGGACACCCATGCTTAACCAACAAGGTTACTGCTACAATCGGACGTGCCACC 1615
Qy 512 LeuGlnGlnGlnTrpTrpGluLeuTrpGlyProGlyThrGluValAlaLysSerTrp 531
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1616 ATGAAGAACCAATGATTTCCTTTCTTTGGGCCAAGTGAACCTGGCTAAAGATTCAATG 1675
Qy 532 TyrAsnArgAsnGluGlyGlySerLysTyrGlyTyrTrpArgArgValTrpTrpLeu 551
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1676 TTCAAACTATACCAGAAAGGCAAGTAGTTATGGCTACTGCAGAAAGAAATGGTACAAG 1735
Qy 552 IleProTrpLysAlaAsnTrpThrMetTrpGlyLysLeuTrpTrpGlnGlyGlySerTrp 571
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1736 ATTCATGTGAACCAACAGAGATGTAATAATGGCAGGTTATTCTGTACCTT----- 1786
Qy 572 AsnLeuProTrpLysGlyA-GIleValThrTrpLeuThrTrpLysThrTrpProGlu 591
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1787 AATAAACCCGGAAGAGAAATAATTGCAACGTGATA-----TATACACC--- 1831
Qy 592 TrpThrSerGlnGluIleGlyMetValAaAsnGlyThrLysTrpGlyTrpAsnLysVal 611
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1832 --ACAGATGAAGATATTGGATGGTTCTTCTCTGGAACAAATGTGGACGTGGAAGGTC 1888
Qy 612 TrpIleAsnAlaGluTrpValTrpIleGluLysAlaThrLysSerThrAsnTrpSerSer 631
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1889 TGCAGCAACGGGCATTTGTTGATGTGCTACAGCCTACTAATCAACCACTGGCTTCTCT 1948
Qy 632 Lys-----TrpLys 634
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1949 CAGATTGATTCTGGAGA 1966
```

Search completed: June 9, 2004, 21:24:00
Job time : 796 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 9, 2004, 20:41:36 ; Search time 7906 Seconds
(without alignments)
4248.780 Million cell updates/sec

Title: US-10-726-148A-15

Perfect score: 4712

Sequence: 1 MLQGLLPVSLLSVAVSAIK.....PPTVKKWNPMTPKWNPXA 775

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US10726148/runat_09062004_161328_5277/app.query.fasta_1.967
-DB=GenEmbl -OPMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10726148 @CGN 1.1 52339 @runat_09062004_161328_5277 -NCPU=6 -ICPU=3
-NO MWAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DBLOP=6 -DELEXT=7

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htg_mus.*
34: em.htg_pln.*
35: em.htg_rod.*
36: em.htg_man.*
37: em.htg_vrt.*
38: em.sy.*
39: em.htgo_hum.*
40: em.htgo_mus.*
41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3128	66.4	2325	6	E44010 Novel prote
2	3128	66.4	3066	6	AJ242015 Homo sapi
3	3120	66.2	2786	9	AF137334 Homo sapi
4	2775.5	58.9	2828	9	AJ242014 Macaca fa
5	2132	45.2	1620	6	E44000 Novel prote
6	2128	45.2	2087	6	AX780064 Sequence
7	2128	45.2	2087	9	AF137335 Homo sapi
8	2098	44.5	1590	6	A61276 Sequence 2
9	2098	44.5	1590	6	AR088819 Sequence
10	2098	44.5	2056	6	AR088818 Sequence
11	2098	44.5	2057	6	A61275 Sequence 1
12	2071.5	44.0	2580	10	BC058782 Mus muscu
13	2071.5	44.0	2638	10	AF163291 Mus muscu
14	2071.5	44.0	2940	10	AF153350 Mus muscu
15	2071.5	44.0	3025	10	AF163290 Mus muscu
16	2071.5	44.0	3110	10	AF163292 Mus muscu
17	1963	41.7	2335	10	AY283187 Rattus no
18	784	16.6	3360	9	MFEAPI X66139 M.fascicula
19	764.5	16.2	2334	6	AX393444 Sequence
20	763	16.2	2334	6	AX393444 Sequence
21	759.5	16.1	2605	6	AX254431 Sequence
22	757	16.1	2326	5	AF450503 Bothrops
23	755.5	16.0	3586	10	RNEAPI X66140 R.norvegicu
24	750.5	15.9	2434	6	AX418414 Sequence
25	750	15.9	2583	9	AF215824 Homo sapi
26	745	15.8	2431	6	AX418412 Sequence
27	742	15.7	2332	5	ACU86634 Agkistrodon
28	733.5	15.6	2045	5	AY101383 Naja moss
29	727	15.4	2368	5	AB074143 Trimeres
30	705	15.0	2248	5	AF051787 Gloydus
31	704	14.9	2238	5	AF063190 Naja naja
32	702	14.9	2187	9	HSDISPRO Y13323 Homo sapien
33	702	14.9	2209	6	AR080771 Sequence
34	702	14.9	2209	6	I79592 Sequence 1
35	699.5	14.8	2290	5	CAU21003 U21003 Crotalus at
36	698.5	14.8	2065	5	EPECHI X78970 E.pyramidum
37	695	14.7	2306	5	AF149788 Bothrops
38	693.5	14.7	2237	5	AB074144 Trimeresu
39	690.5	14.7	2333	5	AF056025 Bothrops
40	690	14.6	2280	6	AR281417 Sequence
41	686.5	14.6	3034	10	MMMSGLYP X13335 M.musculus
42	684	14.5	2331	5	AF490534 Bothrops
43	682.5	14.5	2347	5	AY339162 Macrolope
44	677	14.4	2118	5	BJCARARAC X68251 B.jararaca
45	674.5	14.3	3110	10	BC025584 Mus muscu

ALIGNMENTS

RESULT 1


```
Qy 481 LysSerGlyAsnTrpProTrpTrpArgTrpGlnValAsnGlyTrpProTrpHisHisGly 500
Db 1441 AAATCTGTTATCTGCTGATGATAGATTCACAGTCAATGGCTTCCCTTGGCATCAAGGG 1500
Qy 501 LysGlyHisTrpLeuMetGlyThrTrpProTrpLeuGlnGluGlnTrpThrGluLeuTrp 520
Db 1501 AAGGGCCACTCTGATGGGACATGCCCCACATGACAGGACAGTGCACAGAGCTGGG 1560
Qy 521 GlyProGlyThrGluValAlaTrpLysSerTrpTrpAsnArgAsnGluGlySerHis 540
Db 1561 GGACAGCAACTGAGTGGCAGTAAAGTCATGTTACACAGCAATGAAGTGGTCAAG 1620
Qy 541 TyrGlyTrpTrpArgGlyValTrpTrpThrLeuLeuProTrpLysAlaAsnTrpThrMet 560
Db 1621 TACGGGTACTCTGCGAGAGTGGATGACACATCTCCCTGCAAGCAAAATGATACCAATG 1680
Qy 561 TrpGlyLysLeuTrpTrpGlyGlySerTrpAsnLeuProTrpLysGlyArgIleVal 580
Db 1681 TGTGGAGTGTCTCTCAAGTGGTGGGATATTTGCCCTGCAAGAGCGGATAGTG 1740
Qy 581 ThrTrpLeuThrTrpLysTrpTrpProGluTrpThrSerGlnGluIleGlyMetVal 600
Db 1741 ACTTCTCTGACATGTAAACATTTGATCTGAGACACAAAGTCAAGAAATAGGCATGGTG 1800
Qy 601 AlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpIleAsnAlaGluTrpValTrpIle 620
Db 1801 GCCAATGACATGATGTGGCGATTAACAGGTTTCATTAATGCAAGATGTGGATATT 1860
Qy 621 GluLysAlaTrpLysSerThrAsnTrpSerLysTrpLysGlyHisAlaValTrpTrp 640
Db 1861 GAGAAAGCCTCAAAATCAACAAATGCTCATCTAAGTCAAGGACATGCTGTGTGTGAC 1920
Qy 641 HisGluLeuGlnTrpGlnTrpGluGluGlyTrpIleProProTrpTrpTrpSerSer 660
Db 1921 CATGAGCTCCAGTGTCAATGTGAGAGAGATGGATCCCTCCGACTGCGATGACTCCCTCA 1980
Qy 661 ValValTrpHisTrpSerIleValValGlyValLeuTrpProMetAlaValIleTrpVal 680
Db 1981 GTGGTCTCCACTCTCCATGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT 2040
Qy 681 ValValAlaMetValIleArgHisGlnSerSerArgGluLysGlnLysLysTrpGlnArg 700
Db 2041 GTGGTGTGTATGTTAATCCGGACCCAGAGCTCCAGAGAAAGCAGAGAAAGATCAGAGG 2100
Qy 701 ProLeuSerThrThrGlyThrArgProHisLysGlnLysArgLysProGlnMetValLys 720
Db 2101 CCACATATCTACCTCTGCGACCCAGGCGCCACAAACAGAGAGAGAAACCCCGATGCTAAG 2160
Qy 721 AlaValGlnProGlnGluMetSerGlnMetLysProHisValTrpTrpLeuProValGlu 740
Db 2161 GCTGTTCAACCCCAAGAGATGATGATGAGTGAAGCCCATGTTGATGATCTGCCATAGNA 2220
Qy 741 GlyAsnGluProProAlaSerTrpHisLysTrpThrAsnAlaLeuProProThrValTrp 760
Db 2221 GGCAATGAGCCCCAGCCCTCTTTTATAAAGACACAAACGCACTTCCCTACTGTTTTC 2280
Qy 761 LysTrpAsnProMetSerThrProLysTrpSerAsnProLysAla 775
Db 2281 AAGGATAATCAATGTCTACACCTAAGGACTCAAAATCCAAAGCA 2325

RESULT 2
HSA242015
LOCUS HSA242015 3066 bp mRNA linear PRI 07-JAN-2000
DEFINITION Homo sapiens mRNA for eMDC II protein.
ACCESSION AJ242015
VERSION AJ242015.1 GI:4757043
KEYWORDS eMDC II protein.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
```

```
AUTHORS Jury, J.A., Perry, A.C. and Hall, J.L.
TITLE Identification, sequence analysis and expression of transcripts
encoding a putative metalloproteinase, eMDC II, in human and
macaque epididymis
JOURNAL Mol. Hum. Reprod. 5 (12), 1127-1134 (1999)
MEDLINE 20054911
PUBMED 10587367
REFERENCE 2 (bases 1 to 3066)
AUTHORS Hall, J.L.
TITLE Direct Submission
JOURNAL Submitted (04-MAY-1999) Hall J.L., Biochemistry, University of
Bristol, School of Medical Sciences, University Walk, Bristol, BS8
1TD, UNITED KINGDOM
FEATURES
source
1..3066
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/sex="male"
/tissue_type="epididymis"
/dev_stage="adult"
<1..47
48..2375
/function="putative metalloproteinase"
/codon_start=1
/product="eMDC II protein"
/protein_id="CAB42085.1"
/db_xref="GI:4757044"
/db_xref="GOA:Q9UKQ2"
/translation="MLQGLPVSLLSVAISAIKELPGVKYEVVYPIRLHLKREA
KEPEQOEFTELKVKMTINGKIAVLKLNKNLLAPGYETVYNSTGKEITTSPOIM
DDCYVOGHILANEKVSADSTCTCRGLRGVFSQDQRYFIEPLSPIHEDGOEHALFKYNP
DEKNYDSTCGMDGVLAHDLOQNIAPLAKVLKDKRKQHEKIEYLYVLDNGEFK
RYNEQDSEIRKRVFEMANYNMLYKLNTHVALVGMEIMWDKDKIKITPNASTLENF
SKWRSVLSRRKRHDIAQLITATLGTATGLAFPMSTCPSYSGVGVQDHSNLLRVA
GTMAHEMGHFGMHDDYSCTICVMKALSYIPTDFSSCSRLSYDKPFEDKLS
NCLFNAPLPTDIISTPICNOLVEMGEDCCGTSECTNICCDAKTKIKATFOCALG
BCCEKQCPKAGMCRPAKEDCDLPEMCGNKGSCGDDPFOYNGPCHHGKGLMGT
CPTLQBCQTELPGWGTADVADSCYRNKRGSKYGCRRVDDTLLIPEKANDTWGKLFK
QGSNLLPWKGRIVTFLTKTDPEDTSGEIGVANGTKGDKNGKVCINAECDVIEKAY
KSTNGSSKCGHVAVDLQCEGWIIPDCDDSSVPHFSIVGVLPFMAVIFVIV
AMVTRHQSRRKQKQDQPLSTGTTRPHKQKRPQMKAVQPOQMSQMKPHVYDLPE
GNEPSPFKDTNALPPTVFKDNPMTSPKDSNPKA"
2376..3066
3'UTR
ORIGIN
Alignment Scores:
Pred. No.: 3,05e-188 Length: 3066
Score: 3128.00 Matches: 654
Percent Similarity: 88.13% Conservative: 29
Best Local Similarity: 84.33% Mismatches: 92
Query Match: 66.38% Indels: 0
DB: 9 Gaps: 0
US-10-726-148a-15 (1-775) x HSA242015 (1-3066)
Qy 1 MetLeuGlnGlyLeuLeuProValSerLeuLeuSerValAlaValSerAlaIleLys 20
Db 48 ATGTTGCAAGGTCTCTCCAGTCAGTCTCCCTCTCTGTTGCGAGTAGTCTATAAAA 107
Qy 21 GluLeuProGlyValLysLysTrpGluValValTrpProIleArgLeuHisProLeuHis 40
Db 108 GAACTCCCTGGGTGCAAGAGTATGAGTGGTTTATCTATAGAGCTTCATCCACTGCAT 167
Qy 41 LysArgGluAlaLysGluProGluGlnGlnGluInTrpGluThrGluLeuLysTrpLys 60
Db 168 AAAAGAGAGCGCCAAAGAGCAGCAGCAACAGGAACTTTGAAACTCAATTAAGTATAAA 227
Qy 61 MetThrIleAsnGlyLysIleAlaValLeuTyrLeuLysLysAsnLysAsnLeuAla 80
Db 228 ATGCAAAATTAATGGAATAATTCAGTGTCTTTATTTGAAAAAARCAAGAACTCTCTTGA 287
```

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

QY 81 ProGlyThrGlnThrTyAsnSerThrGlyLysGluLeuThrThrSerProGln 100
Db 288 CCAGGCTACAGCGAAACATATTAATTCACCTGAAAGGAGATCACACAGCCACAA 347
QY 101 LeuMetTrpTrpTrpTyTrpGlnGlyHisLeuLeuGlnLysValSerTrpAlaSer 120
Db 348 ATTATGGATGATTGTATTATCAAGGACATATTCCTTAATGAAAGAGTTCTCGAGCTAGC 407
QY 121 IleSerThrTrpArgGlyLeuArgGlyTyTrpSerGlnGlyTrpGlnArgTyTrpIle 140
Db 408 ATCAGCAGATGAGGGGTCTAAGGGCTACTTCAGTCAGGGGATCAAGATCTTTATT 467
QY 141 GluProLeuSerProIleHisArgTrpGlyGlnGlnHisAlaLeuTrpLysTyAsnPro 160
Db 468 GAACCTTTAAGCCCATACATCGGATGAGCAGGAGCATGCACCTCTTCAAGTATAACCT 527
QY 161 TrpGluLysAsnTyTrpSerThrTrpGlyMetTrpGlyValLeuTrpAlaHisTrpLeu 180
Db 528 GATGAAAGAAATATGACAGCAGCTGTGGATGGATGGTGTGTGGGCTCGAGATTG 587
QY 181 GlnGlnAsnIleAlaLeuProAlaThrLysLeuValLysLeuTyTrpArgLysValGln 200
Db 588 CAGCAGACATTCCTCTACCTGCCACCAACTAGTAATAATGAAAGACAGAGGTTGAG 647
QY 201 GluHisGluLysTyIleGluTyTrpLeuValLeuTrpAsnGlyGluTrpLysArgTy 220
Db 648 GAACATGAGAAATACATAGATAATATTATGCTCTGATGATGATGTTGTGGGCTTAAAGGTAC 707
QY 221 AsnGluAsnGlnTrpGluIleArgLysArgValTrpGluMetAlaAsnTyValAsnMet 240
Db 708 AATGAGANTCAAGATGAGATCAGAAAGGGGTATTTGAGATGGCTAATATTGTCACATG 767
QY 241 LeuTyLysLysLeuAsnThrHisValAlaLeuValGlyMetGluIleTrpTrpLys 260
Db 768 CTTTATAAAGCTCATATCTCATCTGCGCTTAGTGGTATGGAATCTGGATGACAG 827
QY 261 TrpLysIleLysIleThrProAsnAlaSerTrpThrLeuGluAsnTrpSerLysTrpArg 280
Db 828 GATAAGATAAAGATAACCCCAATGCAAGCTTCACCTTGAGAAATTTCTTAAATGGAG 887
QY 281 GlySerValLeuSerArgArgLysArgHisTrpIleAlaGlnLeuIleThrAlaThrGlu 300
Db 888 GGGAGTGTCTCTCAGAAAGACCGTCATGATTTGCTCAGTAAATCAGCAGCAACAGAA 947
QY 301 LeuAlaGlyThrThrValGlyLeuAlaTrpMetSerThrMetTrpSerProTySerVal 320
Db 948 CTGTGCTGAACGACTGTGGGTCTGCATTTATGCTACATGCTCTCTCTATTCTGTT 1007
QY 321 GlyValValGlnTrpHisSerTrpAsnLeuLeuArgValAlaGlyThrMetAlaHisGlu 340
Db 1008 GGGGTGTCTCAGGACCAACGAGATAATCTCTTAGAGTTGAGGGGCAATGGCAGCATGAA 1067
QY 341 MetGlyHisAsnTrpGlyMetTrpHisTrpTyTrpSerTrpLysTrpProSerThrIle 360
Db 1068 ATGGGCCACACTTGGAAATTTTCATGACGACTATTTCTGCAAGTCTCTCTTACATA 1127
QY 361 TrpValMetTrpLysAlaLeuSerTrpTyIleProThrTrpTrpSerSerTrpSerArg 380
Db 1128 TGTGTGATGACAAAGCACTGAGCTCTATATATACCCACACACTTCAGTCTCTCGAGCGGT 1187
QY 381 LeuSerTyTrpLysTrpTrpGluTrpLysLeuSerAsnTrpLeuTrpAsnAlaProLeu 400
Db 1188 CTCAGCTATGACAAAGTTTTTGAAGATAAATATCAAAATGTCCTCTTAAATGCTCCATTG 1247
QY 401 ProThrTrpIleIleSerThrProIleTrpGlyAsnGlnLeuValGluMetGlyGluTrp 420
Db 1248 CCTACAGATATCATATCCACTCCAAATTTGGGACCGGTGTTGGAAATGGGAGAGGAC 1307
QY 421 TrpTrpTrpGlyThrSerGluGluTrpThrAsnIleTrpTrpTrpAlaLysThrTrpLys 440
Db 1308 TGTGATTGTGGACATCTGAGGAATGTACCAATATTTCGTGTCATGCTAAGACATGTAA 1367
QY 441 IleLysAlaThrTrpGlnTrpAlaLeuGlyGluTrpTrpGluLysTrpGlnTrpLysLys 460

Db 1368 ATCAAGCAACTTTTCAATGTGCATTAGGAGAAATGTTGTGAAAAATGCCAATTTAAAAAG 1427
QY 461 AlaGlyMetValTrpArgProAlaLysTrpGluTrpTrpLeuProGluMetTrpAsnGly 480
Db 1428 GCTGGATGTGTGTCAGACCAAGAAAGATGAGTGCAGCTGCCCTGAAATGTGTAATGGT 1487
QY 481 LysSerGlyAsnTrpProTrpTrpArgTrpGlnValAsnGlyTrpProTrpHisHisGly 500
Db 1488 AATCTGTAATTTCTCTGATGATAGATTCCAACTCAATGGCTTCCTTGGCATCACGGG 1547
QY 501 LysGlyHisTrpLeuMetGlyThrTrpProThrLeuGlnGluInTrpThrGluLeuTrp 520
Db 1548 AAGGGCCACTGCTGTATGGGACATGCCACACCTGCAGGAGCAGTCACAGAGCTGTGG 1607
QY 521 GlyProGlyThrGluValAlaTrpLysSerTrpTyAsnArgAsnGluGlyGlySerLys 540
Db 1608 GGACCAAGAACTGAGGTTCAGATAAGTCATGTTACACAGGAATGAAGTGGGTCAAAG 1667
QY 541 TyTGlyTyTrpArgArgValTrpTrpThrLeuIleProTrpLysAlaAsnTrpThrMet 560
Db 1668 TACGGTACTGTCCGAGTGGATGACACACTCATCTCCCTCRAAGCAATGATACCATG 1727
QY 561 TrpGlyLysLeuTrpTrpGlnGlySerTrpAsnLeuProTrpLysGlyArgIleVal 580
Db 1728 TGTGGGAAGTTGTTCTGTCAGGTGGTTCGATTAATTTGCCCTGAAAAGACGATAGTG 1787
QY 581 ThrTrpLeuThrTrpLysThrTrpTrpProGluTrpThrSerGlnGluIleGlyMetVal 600
Db 1788 ACTTCTGACATGTAACAATTTTGATCCTGAGACACAAAGTCAAGAAATAGGCATGTG 1847
QY 601 AlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpIleAsnAlaGluTrpValTrpIle 620
Db 1848 GCCAATGGAACTAAGTGTGGGATACCAAGGTTTGCATTAATGCAGATGTGTGATATT 1907
QY 621 GluLysAlaTyLysSerThrAsnTrpSerSerLysTrpLysGlyHisAlaValTrpTrp 640
Db 1908 GAGAAAGCTTACAAATCAACCAATTTGCTCATCTAAGTGCAGAAAGACATGCTGTGTGTAC 1967
QY 641 HisGluLeuGlnTrpGlnTrpGluGluGlyTrpIleProProTrpTrpTrpTrpSerSer 660
Db 1968 CATGAGCTCCAGTGTCAATGTGAGGAAGGATGGATCCTCCCGACTCGATGACTCTCA 2027
QY 661 ValValTrpHisTrpSerIleValValGlyValLeuTrpProMetAlaValIleTrpVal 680
Db 2028 GTGGTCTTCCACTTCCATTGTGGGTGCTGTTCCTCCATGGCGGTCAATTTTGTG 2087
QY 681 ValValAlaMetValIleArgHisGlnSerSerArgGluLysGlnLysLysTrpGlnArg 700
Db 2088 GTGGTTGCTATGTGTAATCCGGCACCCAGAGCTCCAGAGAAAGACAGAAAGATCAGAGG 2147
QY 701 ProLeuSerThrThrGlyThrArgProHisLysGlnLysArgLysProGlnMetValLys 720
Db 2148 CCACTATCTACCACTGGCCAGCCACACAAACAGAGAGAGAAACCCAGATGTTAAAG 2207
QY 721 AlaValGlnProGlnGluMetSerGlnMetLysProHisValTyTrpLeuProValGlu 740
Db 2208 GCTGTTCAACCCCAAGAGATGATGATGAGCCCATGATGATGATGATGATGATGATGATGAT 2267
QY 741 GlyAsnGluProProAlaSerTrpHisLysTrpThrAsnAlaLeuProProThrValTrp 760
Db 2268 GGCAATGAGCCCCAGCCTCTTTTCATAAAGACACAAACGACTTCCCTCTACTGTTTTC 2327
QY 761 LysTrpAsnProMetSerThrProLysTrpSerAsnProLysAla 775
Db 2328 AAGGATATTCATATGCTACACCTTAGGACTCAATTCACAAAGCA 2372

RESULT 3

AF137334

LOCUS

DEFINITION

ACCESSION

2786 bp mRNA linear PRI 22-SEP-1999
Homo sapiens metalloprotease disintegrin cysteine-rich protein,
transmembrane form mRNA, complete cds.
AF137334

Qy 461 AlaGlyMetValTTPArgProAlaLysTrpGluTrpLeuProGluMetTTPAsnGly 480
 Db 1388 GTGGGATGGTGTGCACACGACCAAGATGAGTGGACCTGCTGAAATGTAAAGGT 1447
 Qy 481 LysSerGlyAsnTrpProTTPArgTrpGlnValAsnGlyTrpProTTPHisHisGly 500
 Db 1448 AAATCTGTAATTCCTCTGATGATAGATTCCAGTCAATGGCTTCCCTTGCATCAOAGG 1507
 Qy 501 LysGlyHisTrpLeuMetGlyThrTrpProThrLeuGlnGluTrpThrGluLeuTrp 520
 Db 1508 AAGGGCCACTGCTTGTATGGGACATGCCCACTGACGAGGAGTGCACAGAGCTGTGG 1567
 Qy 521 GlyProGly 523
 Db 1568 GGACCAAGT 1576

RESULT 12
 BC058782 2580 bp mRNA linear ROD 08-OCT-2003
 DEFINITION Mus musculus a disintegrin and metalloprotease domain 28, mRNA
 [cDNA clone MGC:67767 IMAGR:6310763], complete cds.
 ACCESSION BC058782
 VERSION BC058782.1 GI:37589277
 KEYWORDS MGC.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 2580)
 Strausberg,R.L., Feingold,E.A., Grouse,J.H., Derge,J.G.,
 Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
 Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
 Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
 Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,J.,
 Scapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
 Scheet,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
 Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
 Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
 McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
 Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
 Villaon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
 Fahey,J.J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
 Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
 Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
 Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
 Butlerfield,Y.S., Krzyzanski,M.I., Skalska,U., Smalls,D.E.,
 Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
 Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 22388257
 12477932
 2 (bases 1 to 2580)
 Strausberg,R.
 Direct Submission
 TITLE Submitted (30-SEP-2003) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-rc@mail.nih.gov
 Tissue Procurement: Susan L. Sullivan, PhD.
 cDNA Library Preparation: ResGen, Invitrogen Corp
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural
 Sequencing Center (NISC),
 Gaithersburg, Maryland:
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc_mgc@hgrl.nih.gov
 Akhtar,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
 Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
 Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,

Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Lalic,P., Legaspi,R.,
 Maduro,O.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
 McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
 Turgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
 Young,A., Zhang,L.-H. and Green,E.D.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 127 Row: h Column: 13
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 23956049.
 FEATURES
 source
 1. 2580
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="MGC:67767 IMAGR:6310763"
 /tissue_type="Olfactory epithelium, neonatal mouse,
 C57Bl/6"
 /clone_lib="NIH MGC_129"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 gene
 1. 2580
 /gene="Adam28"
 /note="synonyms: MDC-Lm, MDC-Ls, eMDCII, MDCII, MDC-L,
 D430033C21Rik"
 /db_xref="LocusID:13522"
 /db_xref="MGI:105988"
 28. 2352
 /codon_start=1
 /product="Adam28 protein"
 /protein_id="AAH58782.1"
 /db_xref="GI:37589278"
 /db_xref="LocusID:13522"
 /translation="MQQSLVSVFLLSPVPSVAIKELPKAKKYEYVPIRHLRKR
 ETQPEFETETELRYKVTGKAVLYLKKNNKLAPDYSETYNSGNKVTSPQ
 IMDSYQGHIVNEKVSAAISTCGLRGYISQDEKTYFIPLSENIDEQAHAFKD
 DSNDEKSNCGVDALQGHODVALPATRLIKNDGMVQBPKKYIYYVVDNGE
 PKYNNKLAERKIVLEMANVIMLYNKLDHVALVGVIMTDGDKITPDANTTLE
 NFSKRGMDLKRKHHDIAQLISSTFGSTVGLAFMSMCSPHYSGIVQDHSNYHL
 RVATMAHEMGNLGMIDHLYSCKPCSEVMBOSLRPHMPTDFSSSRVNTKQFLEE
 KLSHLFNPLPSDIITPVCNGLNENEDCGTPEKCTNKCDARTCKIKAGFQC
 ALGECCKOLKPGVVCRAAKDCDLEVCDSKSHCPGRFRVNGSPCQKHGVCCL
 KGKPTTQQCQMMWPGTKYKANTSCYKQNEGTYKYGVCHVNGTHMCKADCMGK
 LFCGSGDLPWKGLTISFLCKLPEDTSQGVDMVANGTKGNTKVCINAEVCVDM
 KTYSSCKSKKHAVCHDLQCKCKEGRAPPDCEASATVPHFSIVGVLPPLAVIF
 VVVAIVTORAKRRQVRQLPSQDKLHNQKCRPKOKVDVQPEMSQMKLHVSD
 LPSEPEPPPPDLITKPNFPPEPPVSLDPAKV"
 247. 552
 /note="Pep M12B proper; Region: Reprolysin family
 propeptide_ This region is the propeptide for members of
 peptidase family M12B. The propeptide contains a sequence
 motif similar to the 'cysteine switch' of the matrixins.
 This motif is found at the C terminus of the alignment but
 is not well aligned"
 /db_xref="CDD:pfam01562"
 643. 1233
 /note="Reprolysin; Region: Reprolysin (M12B) family zinc
 metalloprotease. The members of this family are enzymes
 that cleave peptides. These proteases require zinc for
 catalysis. Members of this family are also known as
 adamalysins. Most members of this family are snake venom
 endopeptidases, but there are also some mammalian proteins
 and fertilin. Fertilin and closely related proteins appear
 to not have some active site residues and may not be
 active enzymes"
 /db_xref="CDD:pfam01421"
 1282. 1506
 /note="DISIN; Region: Homologues of snake disintegrins"
 /db_xref="CDD:smart0050"
 1513. 1896
 /note="ACR; Region: ADAM Cysteine-Rich Domain"
 /db_xref="CDD:smart00608"
 misc_feature
 misc_feature
 misc_feature
 misc_feature

ORIGIN

Alignment Scores:

Pred. No.: 1,14e-121 Length: 2580
 Score: 2071.50 Matches: 449
 Percent Similarity: 74.19% Conservative: 103
 Best Local Similarity: 60.35% Mismatches: 189
 Query Match: 43.96% Indels: 3
 DB: 10 Gaps: 3

US-10-726-148A-15 (1-775) x BC058782 (1-2580)

QY 5 LeuLeuProValSerLeuLeuSer---ValAlaValSerAlaIleLysGluLeuPro 23
 DB CTTCTGTAGTCTCTTCCCTTCTTCTCCAGTCCAGTAAGTCAATAAAGAACTCCCT 102
 QY 24 GlyValLysLysTyrGluValValTyrProIleArgLeuHisProLeuHisLysArgGlu 43
 DB AAAGCCCAAGAAATATGAAGTGGTTATCCCAATAGACTTCATCCATTTGCGTAAAGAGAG 162
 QY 44 AlaLysGluProGluGlnGlnGlnTyrGluLeuLysTyrLysMetThrIle 63
 DB ACCCAAGAGCCAGAGCCAAAGAAACATTTGAACTGAGCTTAGGTHCAAAATGACAGTA 222
 QY 64 AsnGlyLysIleAlaValLeuLysLeuLysLysAsnLysAsnLeuLeuAlaProGlyTyr 83
 DB AATGGAAGGTGCTGTGTATCTCAAGAGAAACAAACAAAGCTCCTTGGCCCTGACTAC 282
 QY 84 ThrGluThrTyrTyrAsnSerThrGlyLysGluIleThrThrSerProGlnIleMetTyr 103
 DB TCGGAACACATACATATAATTCGTTGGAACCAAGGTCCACCAAGCCGCAATCATGGAT 342
 QY 104 TrrTrrTrrTrrGlnGlnHisIleLeuAsnGluLysValSerTrrAlaSerIleSerThr 123
 DB ASCTGTACTACCAAGACACATGCTAAATGAGAAAGTTTCTGCAGCCACATCAGCACC 402
 QY 124 TrpArgGlyLeuArgGlyTyrTrrSerGlnGlyTrrPglNArgTrrPllleGluProLeu 143
 DB TGTCAAGGACTACGGGTTCATCATCAGTCAAGGAGTGAATTTTATCGAACCTTTG 462
 QY 144 SerProIleHisArgTrrPglGlyGlnGluHisAlaLeuTrrLysTyrAsnProTrrGluLys 163
 DB AGCTCGAGAACTTGGATGAAACAGGCACATGCTCTTCAAGGACGACTCCCAATGAAGAC 522
 QY 164 AsnTrrTrrSerThrTrrPglMetTrrPglValLeuTrrPalaHisTrrLeuGlnGlnAsn 183
 DB CAGGAGAGAGCAACTGTGTGTGGATGATCGCTATGGCTCCAAGGCTGCATCAGAC 582
 QY 184 IleAlaLeuProAlaThrLysLeuValLysLeuLysTrrPargLysValGlnGlnHisGlu 203
 DB GTGGCCCTCCCTCCACCGAGTTGATTAAGTTGAATGATGGGATGGTTCAGAAACCTTAAG 642
 QY 204 LysTrrIleGluTrrTrrLeuValLeuTrrPalaHisTrrLysTrrPglNArgTrrPllleGlu 223
 DB AATATACATAGATATATATGCTGCTCGATTAAGTGAATGATGGGATGGTTCAGAAACCTTAAG 702
 QY 224 GlnTrrPglLysArgLysArgValTrrPglMetAlaAsnTrrValAsnMetLeuTrrLys 243
 DB CTTGCTGAATACAGAAAGATAGTCTTGAGATGGCCCAATTACATCAACATGCTTTACAT 762
 QY 244 LysLeuAsnThrHisValAlaLeuValGlyMetGluIleTrrPglNArgTrrPllleGlu 263
 DB AAGCTTGAATCCCAACGCTGAGTCTGATTAAGTGAATGATGGGATGGTTCAGAAACCTTAAG 822
 QY 264 LysIleThrProAsnAlaSerTrrTrrLeuGlnAsnTrrSerLysTrrPargLysSerVal 283
 DB AAGATACACAGATGCCAACACCCCTCGAAACATCTTCTAAGTGGAGGGAAATGAT 882
 QY 284 LeuSerArgArgLysArgHisTrrPllleAlaGlnLeuIleThrAlaThrGluLeuAlaGly 303
 DB CTGCTAAACGAAAGCATCATGATATGSCCAGCTAATCTCATCAACAGACTTTTCTGGA 942
 QY 304 ThrThrValGlyLeuAlaTrrMetSerThrMetTrrSerProTrr---SerValGlyVal 322

DB TCAACAGTGGGTCTAGCCTTCATGTGTCGATGTGTTCACTTACCATTTCTGTGGCAT 1002
 QY 323 ValGlnTrrPalaHisSerTrrPalaHisValAlaGlyThrMetAlaHisGluMetGly 342
 DB GTTCAGGACCAACAGTAACTACCATCTTCGAGTCGAGCAACAAATGGCTCATGAATGGGT 1062
 QY 343 HisAsnTrrPglMetTrrPalaHisTrrTrrSerTrrLysTrrPrrProSerThrIleTrrPVal 362
 DB CACAACTTGGCATGATTCATGACTCTTGAGCTGCAAGTGTCCATCTCGAAGTCTGTGTA 1122
 QY 363 MetTrrLysAlaLeuSerTrrTrrIleProThrTrrTrrSerTrrSerTrrSerArgLeuSer 382
 DB ATGGAGCAGTCACTAAGTTCATATGCTACAGACTTCAGCTCTGTCAGTCTGTGTCAAT 1182
 QY 383 TrrTrrLysTrrTrrPglTrrLysLeuSerAsnTrrPglTrrPalaHisTrrPalaHisProLeuProThr 402
 DB TACAAACAGTTTCTTGAGAAATAATTCGCAATTCCTCTTAAATAGCCCATTCGCACTCA 1242
 QY 403 TrrIleIleSerThrProIleTrrPglValAsnGlnLeuValGluMetGlyGluTrrTrrTrr 422
 DB GATATCATATCCACCCAGTCTGTGGCAACAGTTGTGGAAATGAATGAGAGCTGTGAC 1302
 QY 423 TrrPglTrrSerGluGluTrrTrrAsnIleTrrTrrTrrPalaHisTrrTrrLysIleLys 442
 DB TGTGGCACACCAAGAGTGTACTTAACAATGCTGTGATGCAAGCACCTGTAAATATAA 1362
 QY 443 AlaThrTrrPglNTrpAlaLeuGlyGluTrrPglLysTrrPglLysTrrPglLysLysGly 462
 DB GCAGTTTCCAGTGTCCCTGGGGAAATGCTGAGAAATCCCACTTAAATAAACCTGGG 1422
 QY 463 MetValTrrPargProAlaLysTrrPglTrrPglLeuProGluMetTrrPalaHisLysSer 482
 DB GTTGTGTCAGAGCAGCAAGAGAGTGTGATCTGCTGAGTGTGAGTGTGAGTAAATCC 1482
 QY 483 GlyAsnTrrProTrrTrrPargTrrPglNValAsnGlyTrrProTrrPalaHisLysGly 502
 DB AGCCACTGCCAGGTGACAGATTCAGATTCATAGTCCCTCCCTTGCCTCAAAATGGCATGGT 1542
 QY 503 HisTrrPalaMetGlyTrrTrrProTrrPglNValAsnGlyTrrProTrrPalaHisLysGly 522
 DB TACTGCTTGAAGGGCAAAATGTCCCACTGTCAGCAGTGCATGAGCATGTGGGTGCTCA 1602
 QY 523 GlyThrGluValAlaTrrLysSerTrrTrrAsnArgAsnGluGlyGlySerLysTrrGly 542
 DB GGAACCAAGTTCGAATATCATCATGTTCAAGCAGATGAGTGTGGGACAAAGTACGGA 1662
 QY 543 TrrTrrPargValTrrTrrTrrPglLysTrrPalaHisTrrPalaHisTrrPalaHisGly 562
 DB TACTGTATGTGAGAGATGGCACACATGCTCCCTGCAAAAGCAAAAGATGCTGTGGG 1722
 QY 563 LysLeuTrrTrrPglGlyGlySerTrrPalaHisTrrPglLysTrrPglLysTrrPglLys 582
 DB AATTTGTTGTCGAGGCGGATCAGGTGATTTGCTTGGGAGGAGGACTTACCATATCTTC 1782
 QY 583 LeuThrTrrLysTrrTrrTrrPglTrrPglTrrPglTrrPglTrrPglTrrPglTrrPgl 602
 DB CTGACATGTAATATTATTGATCTCTGAAGACACAAGTCAAGGAGTGAAGTGGTGGCCAAT 1842
 QY 603 GlyThrLysTrrPglTrrPalaHisTrrPglLysTrrPglLysTrrPglLysTrrPglLys 622
 DB GGAACCAAGTGTGGAATCAAGAGTGTGATTAATGCTGAGTGTGGGATGAGGAGGAG 1902
 QY 623 AlaTrrLysSerThrAsnTrrPalaHisTrrPglLysTrrPglLysTrrPglLysTrrPglLys 642
 DB ACTTACAGTACCAACTCTCTCTAAGTGCAGGGGACGAGTGTGGTGGTGGTGGTGGT 1962
 QY 643 LeuGlnTrrPglNTrpGluGlyTrrPglLysTrrPglTrrPglTrrPglTrrPglTrrPgl 662
 DB CTTGAGTGTGAGTCAAGGAGGATGGGCTCTCTGACTGCGAGAAATTCAGCCACAGTC 2022
 QY 663 TrrPalaSerIleValValGlyValLeuTrrPalaHisTrrPalaHisTrrPalaHis 682

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 3, 2004, 23:36:21 ; Search time 801 Seconds
(without alignments)
4413.916 Million cell updates/sec

Title: US-10-726-148A-15

Perfect score: 4712

Sequence: 1 MLQGLIPVSLLSVAVSAIK.....PPTVKNPMTPTKWSNPKA 775

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Ddelop 6.0 , Delext 7.0

Searched: 2995936 seqs, 2280998010 residues

Total number of hits satisfying chosen parameters: 5991872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MCDEL-frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10726148/runat_09062004_161331_5374/app_query.fasta_1.967
-DB=Published Applications NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=C -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cgi -LIST=45 -DLOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10726148@cgn_1_1_549@runat_09062004_161331_5374
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELCP=6 -DELEXT=7

Database : Published Applications NA.*

1:	/cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2:	/cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3:	/cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4:	/cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5:	/cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6:	/cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7:	/cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8:	/cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9:	/cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10:	/cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11:	/cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2.*
14:	/cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
15:	/cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
16:	/cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
17:	/cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
18:	/cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq2.*
19:	/cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
------------	-------------	-------	--------	-------	-------------

RESULT 1
US-10-240-425-1097
; Sequence 1097, Application US/10240425
; Publication No. US20040033502A1
; GENERAL INFORMATION:
; APPLICANT: Williams, Amanda
; APPLICANT: Boland, Joseph F.
; APPLICANT: Lord, Reginald V.
; APPLICANT: Alvarez, Chris
; APPLICANT: Wetzel, Jon C.
; APPLICANT: Scherf, Uwe
; APPLICANT: Vockley, Joseph G.
; TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue
; FILE REFERENCE: 44921-5026
; CURRENT APPLICATION NUMBER: US/10/240,425
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: PCT/US01/09847
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 60/193,446
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 1588
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1097

ALIGNMENTS

1	3128	66.4	3066	13	US-10-240-425-1097	Sequence 1097, Ap
2	3128	66.4	3179	13	US-10-276-774-949	Sequence 949, App
3	1610	34.2	2047	9	US-09-955-504-4	Sequence 4, Appli
4	1610	34.2	2047	14	US-10-125-470-4	Sequence 4, Appli
5	1610	34.2	2047	14	US-10-125-452-4	Sequence 4, Appli
6	1610	34.2	2047	15	US-10-106-698-417	Sequence 417, App
7	854	18.1	1652	9	US-09-925-299-241	Sequence 241, App
8	854	18.1	1652	10	US-09-925-299-241	Sequence 241, App
9	820	17.4	1704	9	US-09-955-504-5	Sequence 5, Appli
10	820	17.4	1704	14	US-10-125-470-5	Sequence 5, Appli
11	820	17.4	1704	14	US-10-125-452-5	Sequence 5, Appli
12	763	16.2	2334	9	US-09-921-823-16	Sequence 16, Appli
13	755	16.0	2634	13	US-10-232-972B-3	Sequence 3, Appli
14	755	16.0	2634	13	US-10-232-972B-1	Sequence 3, Appli
15	750.5	15.9	2434	9	US-09-908-193-9	Sequence 9, Appli
16	745	15.8	2431	9	US-09-908-193-7	Sequence 7, Appli
17	739	15.7	2335	9	US-09-996-620-9	Sequence 9, Appli
18	736	15.6	2359	9	US-09-996-620-17	Sequence 17, App
19	736	15.6	2411	9	US-09-764-898-47	Sequence 47, Appli
20	731.5	15.5	2050	9	US-09-996-620-5	Sequence 5, Appli
21	702	14.9	1462	14	US-10-114-893-7	Sequence 7, Appli
22	702	14.9	2187	13	US-10-342-887-1497	Sequence 1497, Ap
23	702	14.9	2187	13	US-10-172-118-1497	Sequence 1497, Ap
24	690	14.6	2280	15	US-10-349-806-1	Sequence 1, Appli
25	677.5	14.4	2309	9	US-09-996-620-13	Sequence 13, Appli
26	671.5	14.3	2297	9	US-09-996-620-7	Sequence 7, Appli
27	662	14.0	2288	9	US-09-996-620-11	Sequence 11, Appli
28	648.5	13.8	3227	15	US-10-226-844-2	Sequence 2, Appli
29	648.5	13.8	3236	9	US-09-954-456-43	Sequence 43, Appli
30	648.5	13.8	3236	13	US-10-240-425-1213	Sequence 213, Ap
31	648.5	13.8	3236	13	US-10-342-887-560	Sequence 560, App
32	648.5	13.8	3236	13	US-10-172-118-560	Sequence 57, Appli
33	648.5	13.8	3236	15	US-10-211-858-57	Sequence 57, Appli
34	648.5	13.8	3236	15	US-10-210-951-57	Sequence 57, Appli
35	648.5	13.8	3236	15	US-10-211-884-57	Sequence 1, Appli
36	589.5	12.5	2029	15	US-10-439-532-1	Sequence 1, Appli
37	548	11.6	1833	9	US-09-921-823-18	Sequence 18, Appli
38	548	11.6	1833	9	US-09-921-823-21	Sequence 21, Appli
39	545	11.6	2757	9	US-09-955-504-2	Sequence 2, Appli
40	545	11.6	2757	14	US-10-020-733-3	Sequence 3, Appli
41	545	11.6	2757	14	US-10-125-470-2	Sequence 2, Appli
42	545	11.6	2757	14	US-10-125-452-2	Sequence 2, Appli
43	545	11.6	2757	16	US-10-274-639-39	Sequence 39, Appli
44	545	11.6	2757	17	US-10-333-574-39	Sequence 39, Appli
45	545	11.6	2868	14	US-10-020-733-7	Sequence 7, Appli

LENGTH: 3066
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20040033502A1 AJ242015
US-10-240-425-1097

Alignment Scores:

Pred. No.: 1,33e-272 Length: 3066
Score: 3128.00 Matches: 654
Percent Similarity: 88.13% Conservative: 29
Best Local Similarity: 84.39% Mismatches: 92
Query Match: 66.38% Indels: 0
DB: 13 Gaps: 0

US-10-726-148a-15 (1-775) x US-10-240-425-1097 (1-3066)

```
QY 1 MetLeuGlnGlyLeuLeuProValSerLeuLeuSerValAlaValSerAlaIleLys 20
DB 48 ATGTTGCAAGGTCCTGCGCAGTCAGTCTCCTCCTCTCTGTTGAGTAAGTGTATATAA 107
QY 21 GluLeuProGlyValLysLysTyrGluValValTyrProIleArgLeuHisProLeuHis 40
DB 108 GAATCCCTCGGGGGAAGAGATATGAAGTGGTTTATCCTATTAAGACTTCATCCACTGCAT 167
QY 41 LysArgGluAlaLysGluProGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 60
DB 168 AAGAAGAGGCGCAAGAGCCAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAGCAAGAG 227
QY 61 MetThrIleAsnGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 80
DB 228 ATGCAATTAATGAAATTTGCAAGTTCCTTATTTGAAATTTGAAATTTGAAATTTGAA 287
QY 81 ProGlyTyrThrGluThrTyrTyrAsnSerThrGlyLysGluIleThrThrSerProGln 100
DB 288 CAGGCTACACGGAAACATATTAATTAATTCCTCGAAGGAGATCACACCAAGCCCAAA 347
QY 101 IleMetTrpTrpTyrTyrGlnGlyHisIleLeuAsnGluLysValSerTrpAlaSer 120
DB 348 ATATGGATGATGTTATATATCAAGGACATATTTCTTAATGAAGAGTTCTGACGCTAGC 407
QY 121 IleSerThrTrpArgGlyLeuArgGlyTyrTrpSerGlnGlyTrpGlnArgTyrTrpIle 140
DB 408 ATCAGCATGTAGGGGCTAAGGGGCTACTTCTAGTCAGGGGGATCAAGATATCTTTAT 467
QY 141 GluProLeuSerProIleHisArgTrpGlyGlnGlnHisAlaLeuTrpLysTyrAsnPro 160
DB 468 GAACCTTTAGCCCATACATCGGGATGACAGGAGCATGCCTCTTCAAGTATAACCCCT 527
QY 161 TrpGlyLysAsnTyrTrpSerThrTrpGlyMetTrpGlyValLeuTrpAlaHisTrpLeu 180
DB 528 GATGAAAGAAATATGACAGCACCTGTGGGATGATGTTGTGGGCCACGATTTG 587
QY 181 GlnGlnAsnIleAlaLeuProAlaThrLysLeuValLysLysLysLysLysLysLysLys 200
DB 588 CAGCAGAACATTCGCCCTACCTCCACCAAACTAGTAAAAATTAAGAGAGAGGTTTCAAG 647
QY 201 GluHisGluLysTyrIleGluTyrTyrLeuValLeuTrpAsnGlyGluTrpLysArgTyr 220
DB 648 GAACATGAGAAATACATAGAAATATATTTGGTCTCGATATGTTGGATTTTAAAGGTAC 707
QY 221 AsnGluAsnGlnTrpGluIleArgLysArgValTrpGluMetAlaAsnTyrValAsnMet 240
DB 708 AATGAGAATCAAGATGAGATCAGAAAGAGGTTATTCAGATGGCTAATTAATGTAACATG 767
QY 241 LeuTyrLysLysLeuAsnThrHisValAlaLeuValGlyMetGluIleTrpThrTrpLys 260
DB 768 CTTTATAAAGAGCTCAATACTACTCATGTGCGCTTACTTGGTATGGAATCTGAGCTGCAAG 827
QY 261 TrpLysIleLysIleThrProAsnAlaSerTrpThrLeuGluAsnTrpSerLysTrpArg 280
DB 828 GATAGATTAAGATTAACCCCAATGCAAGCTTCACCTTGGAGAAATTTTCTTAAATGGAGG 887
```

```
QY 281 GlySerValLeuSerArgArgLysArgHisTrpIleAlaGlnLeuIleThrAlaThrGlu 300
DB 888 GGGAGTGTCTCTCAAGAGAAAGCGTCATGATTTGCTCAGTTAATCACACCAACAGAA 947
QY 301 LeuAlaGlyThrThrValGlyLeuAlaTrpMetSerThrMetTrpSerProTyrSerVal 320
DB 948 CTTGCTGGAACGACTGTGGGTCTTGCAATTATGCTACAAATGTGCTCTCTTATCTGTT 1007
QY 321 GlyValValGlnTrpHisSerTrpAsnLeuLeuArgValAlaGlyThrMetAlaHisGlu 340
DB 1008 GGCCTGTGTCAGGACCAAGCAATAATCTTCTAGAGTTGCGAGGACAAATGCGCATGAA 1067
QY 341 MetGlyHisAsnTyrGlyMetTyrHisTrpTyrTyrSerTrpLysTrpProSerThrIle 360
DB 1068 ATGGGCCCAACTTGGAGATGTTTCATGACCACTATTTCTGCAAGTGTCTCTTCAACA 1127
QY 361 TrpValMetTrpLysAlaLeuSerTyrIleProThrTrpTrpSerSerTrpSerArg 380
DB 1128 TGTGTGATGGACAAAGCACTGAGCTTCATATATCCACAGACTTCAGTTCTCTGCAAGCT 1187
QY 381 LeuSerTyrTrpLysTrpTrpGluTrpLysLeuSerAsnTrpLeuTrpAsnAlaProLeu 400
DB 1188 CTCAGCTATGACAAAGTTTTTTGAAGATTAATATCAAAATGCTCTTTAATGCTCCATG 1247
QY 401 ProThrTrpIleIleSerThrProIleTrpGlyAsnGlnLeuValGluMetGlyGluTrp 420
DB 1248 CCTACAGATATCATATCCATCCATTTGTGGGACCAAGTTGTTGGAAATGGGAGAGAC 1307
QY 421 TrpTrpTrpGlyThrSerGluGluTrpThrAsnIleTrpTrpTrpAlaLysThrTrpLys 440
DB 1308 TGTGATTTGTGGGACATCTGAGGAATGTACCAATATTTGCTGTGATGCTTAAGACATG 1367
QY 441 IleLysAlaThrTrpGlnTrpAlaLeuGlyGluTrpTrpGluLysTrpGluTrpLysLys 460
DB 1368 ATCAAGCACTTTTCAATGTGCATAGAGAAATGTTGTGAAATATGCCAATTTTAAAG 1427
QY 461 AlaGlyMetValTrpArgProAlaLysTrpGluTrpTrpLeuProGluMetTrpAsnGly 480
DB 1428 GCTGGATGTTGTGCAGACCAAGCAAGATGAGTCCGACCTTCCTGAAATGTGTAATGGT 1487
QY 481 LysSerGlyAsnTrpProTrpTrpArgTrpGlnValAsnGlyTrpProTyrHisGly 500
DB 1488 AAATCTGTGTAATTTCTCTGATGATAGATTTCCAAGTCAATGCTCTTCCCTTGCATC 1547
QY 501 LysGlyHisTrpLeuMetGlyThrTrpProThrLeuGlnGlnTrpThrGluLeuTrp 520
DB 1548 AGGGCCACTGCTTGTATGGGACATGCCACACTGCAGGAGCAGTGCCACAGAGCTGTGG 1607
QY 521 GlyProGlyThrGluValAlaTrpLysSerTrpTyrAsnArgAsnGluGlyGlySerLys 540
DB 1608 GGACCAAGCACTGAGTTGCAGATAAGTCANGTTACAACAGCAATGAAGTGGGTCAAAG 1667
QY 541 TyrGlyTyrTrpArgArgValTrpTrpThrLeuIleProTrpLysAlaAsnTrpThrMet 560
DB 1668 TACGGTACTGTCAGAGTGGATGACACACTCATTCCTGCAAGCAAAATGATACCAATG 1727
QY 561 TrpGlyLysLeuTrpTrpGlnGlySerTrpAsnLeuProTrpLysGlyArgIleVal 580
DB 1728 TGTGGAGTTGTCTGTCAAGTGGGTGCGATTAATTTGCCCTGGAAAGGACGAGTAGTG 1787
QY 581 ThrTrpLeuThrTrpLysThrTrpTrpProGluTrpThrSerGlnGluIleGlyMetVal 600
DB 1788 ACTTTCCTGACATGTAATAACATTTGATCCTGAAGACACCAAGTCAAGAAATAGGCATG 1847
QY 601 AlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpIleAsnAlaGluTrpValTrpIle 620
DB 1848 GCCAATGGAACTAAGTGTGCGATAAAGGTTTTCATTAATGACAGATGTTGTGGATATT 1907
QY 621 GluLysAlaTyrLysSerThrAsnTrpSerLysTrpLysGlyHisAlaValTrpTrp 640
DB 1908 GAGAAAGCCCTACAAATCAACCAATTTGCTCATCTAAGTGCMAAGGACATGCTGTGTGAC 1967
QY 641 HisGluLeuGlnTrpGlnTrpGluGluGlyTrpIleProTrpTrpTrpTrpTrpSerSer 660
```


Db 1308 TGTGATTGTGGGACATCTGAGGAATGTACCAATATTGTGTGTGATGCTAAGACATGTAAA 1367
Qy 441 IleLysAlaThrTrpGlnTrpAlaLeuGlyGluTrpTrpGluLysTrpGlnTrpLysLys 460
Db 1368 ATCAAGCAACTTTTCAATGTGATAGGAGATGTTGTGAAATAATCCCAATTTTAAAG 1427
Qy 461 AlaGlyMetValTrpArgProAlaLysTrpGluTrpTrpLeuProGluMetTrpAsnGly 480
Db 1428 GCTGGGATGTGTGCACACAGCAAGATGATGCGCACTGCTGAAATGTGTAAAGT 1487
Qy 481 LysSerGlyAsnTrpProTrpTrpArgTrpGlnValAsnGlyTrpProTrpHisHisGly 500
Db 1488 AAATCTGGTAATGTCTCGATGATAGATTCCAAGTCAATGGCTCCCTTGCCATCAAGG 1547
Qy 501 LysGlyHisTrpLeuMetGlyTrpProTrpLeuGlnGluTrpTrpGluLeuTrp 520
Db 1548 AAGGGCACTGCTGTGATGGGAGATGCCCACTCGAGAGCAGTGCAAGAGCTGTGG 1607
Qy 521 GlyProGlyThrGluValAlaTrpLysSerTrpTrpAsnArgAsnGluGlyGlySerLys 540
Db 1608 GGACCAAGGAATCTGAGTTCAGATAAGTCATGTTACACAGGAATGAGGTGGTCAAG 1667
Qy 541 TyrGlyTrpTrpArgValTrpTrpTrpLeuLeuProTrpLysAlaAsnTrpThrMet 560
Db 1668 TACGGGTACTGTGCGAGTGGATGACACACTCATTCCTGCAAGCAAAATGATACCATG 1727
Qy 561 TrpGlyLysLeuTrpTrpGlnGlyGlySerTrpAsnLeuProTrpLysGlyArgLeuVal 580
Db 1728 TGTGGGAAGTGTCTGCAAGTGGTGGATTAATTTGCCCTGGAAAGCAGGATGTG 1787
Qy 581 ThrTrpLeuThrTrpLysThrTrpTrpTrpGluTrpTrpSerGlnGluLeuGlyMetVal 600
Db 1788 ACTTCTCGACATGTAAACATTTGATCTCTGGAAGACCAAGTCAAGAAATAGGATGGTG 1847
Qy 601 AlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpIleAsnAlaGluTrpValIle 620
Db 1848 GCCAATGGCACTAAGTGTGGCGATAACAAGTGTTCATTAATGCAGAAATGTGGATATT 1907
Qy 621 GluLysAlaTrpLysSerThrAsnTrpSerSerLysTrpLysGlyHisAlaValTrpTrp 640
Db 1908 GAGAAAGCTTACAAATCAACCAATGCTCATCTAAGTGCAGAAAGACATGCTGTGTGAC 1967
Qy 641 HisGluLeuGlnTrpGlnTrpGluGlyTrpLysProTrpTrpTrpTrpSerSer 660
Db 1968 CATGAGCTCCAGTGTCAATGTGAGGAAGATGGATCCCTCCGACTGCGATGACTCTCA 2027
Qy 661 ValValTrpHisTrpSerIleValValGlyValLeuTrpProMetAlaValIleTrpVal 680
Db 2028 GTGGTCTTCACCTTCCTCAFTGGTGGGGTGGTCTGTCCTCCCAATGGCGGTCAATTTTGTG 2087
Qy 681 ValValAlaMetValIleArgHisGlnSerSerArgGluLysGlnLysLysTrpGlnArg 700
Db 2088 GTGGTGTCTATGTATTCGGGACCCAGCTCCAGGAAAGACAGAAAGATCAGAGG 2147
Qy 701 ProLeuSerThrThrGlyThrArgProHisLysGlnLysArgLysProGlnMetValLys 720
Db 2148 CCACTATCTACCACTGGCACCGGCCACACAAACAGAGAGGAAACCCAGATGGTAAAG 2207
Qy 721 AlaValGlnProGlnGluMetSerGlnMetLysProHisValTrpTrpLeuProValGlu 740
Db 2238 GCTGTTCAACCCCAAGAGATGAGTCAGATGAAGCCCCATGTGATGATCTGCCAGTAGAA 2267
Qy 741 GlyAsnGluProProAlaSerTrpHisLysTrpTrpAsnAlaLeuProProThrValTrp 760
Db 2268 GGCAATGAGCCCCAGCCTCTTTTCATAAAGACACAAACGCACTTCCCTTACTGTTC 2327
Qy 761 LysTrpAsnProMetSerThrProLysTrpSerAsnProLysAla 775
Db 2328 AAGGATAATCCAATGTCTACACCTAAGGACTCAATCCAAATCCAAAGCA 2372

RESULT 3

US-09-955-504-4

; Sequence 4, Application US/09955504

; Publication No. US20020182702A1
; GENERAL INFORMATION:
; APPLICANT: Rubec et al.
; TITLE OF INVENTION: ADAM Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: 77006P2
; CURRENT APPLICATION NUMBER: US/09/955,504
; PRIOR FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: 60/234,222
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 09/712,907
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: PCT/US00/14308
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/178,717
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/142,930
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 60/136,388
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 2047
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-955-504-4

Alignment Scores:
Pred. No.: 1,24e-135 Length: 2047
Score: 1610.00 Matches: 345
Percent Similarity: 83.72% Conservative: 20
Best Local Similarity: 79.13% Mismatches: 71
Query Match: 34.17% Indels: 0
DB: 9 Gaps: 0

US-10-726-148A-15 (1-775) x US-09-955-504-4 (1-2047)

Qy 340 GluMetGlyHisAsnTrpGlyMetTrpHisTrpTrpTrpLysTrpProSerThr 359
Db 11 GAAATGGGCCCAACTTTGGAATGTTTCATGACGACATTTCTTGCAGATGCTCTTACA 70
Qy 360 IleTrpValMetTrpLysAlaLeuSerTrpTrpLysTrpTrpTrpSerTrpSer 379
Db 71 ATATGTGTGATGACCAAGCACTGAGCTTCATATATACCCACAGACTTCAGTTCCTGCAGC 130
Qy 380 ArgLeuSerTrpTrpLysTrpTrpGluTrpLysLeuSerAsnTrpLeuTrpAsnAlaPro 399
Db 131 CGTCTCAGCTATGACCAAGTTTTGAAGATAAATATCAAAATGCTCTTAAATGCTCCA 190
Qy 400 LeuProThrTrpIleLeuSerThrProIleTrpGlyAsnGlnLeuValGluMetGlyGlu 419
Db 191 TTGCCTACAGATATCATATCCACTCCCAATTTGTGGNACCACTGGTGGAAATGGGAGAG 250
Qy 420 TrpTrpTrpTrpGlyThrSerGluGluTrpTrpAsnIleTrpTrpTrpAlaLysThrTrp 439
Db 251 GACTGTGATGTGGGACATCTGAGGAATGTACCAATATTGCTGTGATGCTAAGACATGT 310
Qy 440 LysIleLysAlaThrTrpGlnTrpAlaLeuGlyGluTrpTrpGluLysTrpGlnTrpLys 459
Db 311 AAATCAAGCAACTTTTCAATGTGATAGGAGAAATGTTGTGAAAATAATGCCAATTTAAA 370
Qy 460 LysAlaGlyMetValTrpArgProAlaLysTrpGluTrpTrpLeuProGluMetTrpAsn 479
Db 371 AAGGCTGGGATGTGTGCGACAGCAAGAGATGAGTGGCACTGCTGCTGAAATGTGTAAT 430
Qy 480 GlyLysSerGlyAsnTrpProTrpTrpArgTrpGlnValAsnGlyTrpProTrpHisHis 499
Db 431 GGTAAATCTGGTAATGTCTCTGATGATAGATTCCAAGTCAATGGCTTCCCTTGCATCAC 490
Qy 500 GlyLysLysHisTrpLeuMetGlyThrTrpProThrLeuGlnGluGlnTrpTrpGluLeu 519
Db 491 GGGAAAGGCACTGTGTGATGGGAGCATGCCCAACACATGCGAGGAGGAGTGCACAGAGCTG 550

Qy	520	TrpGlyProGlyThrGluValAlaIrrPlySerTrpTyrAsnA-gAsnGluGlyGlySer	539
Db	551	TGGGGACAGAACTGAGGTTGCAGATAAAGTCATGTTACAACAGCAATGAAGTGGGTCA	610
Qy	540	LysTyrGlyTyrTrpArgArgValTrpThrLeuIleProTfPlyAlaAsnTrpThr	559
Db	611	AAGTACGGGTACTGTCGAGAGTGGATGACACATCATTCCTCGCAAGCAAAATGATACC	670
Qy	560	MetTrpGlyLysLeuTrpTrpGlnGlySerTrpAsnLeuProTfPlyGlyArgile	579
Db	671	ATGNGTGGGAATGTTCTGTCAGGTGGGTCCGATAATTTCCTCGAAGAGCAGGATA	730
Qy	580	ValThrTrpLeuThrTrpLysThrTrpTrpProGluTrpThrSerGlnIleGlyMet	599
Db	731	GTGACTTTCCTGACATGTAAACATTTGATCCTCAAGACACAAGTCAAGAAATAGGCATG	790
Qy	600	ValAlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpIleAsnAlaGluTrpValTrp	619
Db	791	GTGGCCAAATGGAACTAAGTGGCGATAACAAGGTTTGCAATTAAAGCAGAAATGTGTGAT	850
Qy	620	IleGluLysAlaTyrLysSerThrAsnTrpSerSerLysTrpLysGlyHisAlaValTrp	639
Db	851	ATTGAGAAGGCTACAAATCAACCAATTGCTCATCTAAGTCCAAAGGACATGCTGTGTGT	910
Qy	640	TrpHisGlnLeuGlnTrpGlnTrpGluGluGlyTrpIleProProTfTrpTrpTrpSer	659
Db	911	GACCATCAGGCTCCAGTGTCAATGTGAGGAAGATGGATCCCTCCGACATGCGATCACTCC	970
Qy	660	SerValValTrpHisTrpSerIleValValGlyValLeuTrpProMetAlaValIleTrp	679
Db	971	TCAGTGTCTTCACACTTCTCCATTGTGGTGGGTGCTGTTCCTCAATGGCGGTCAATTTT	1030
Qy	680	ValValValAlaMetValIleArgHisGlnSerSerArgGluLysGlnLysLysTrpGln	699
Db	1031	GTGGTGTGTGTAATGGTAAATCCGCGACACAGACTCCAGAGAAAGCGAAGAAATCAG	1090
Qy	700	ArgProLeuSerThrThrGlyThrArgProHisLysGlnLysArgLysProGlnMetVal	719
Db	1091	AGGCCATATCTATCACTCGGCACCGGCCACACAAACAGAAAGAGAAACCCAGATGGTA	1150
Qy	720	LysAlaValGlnProGlnGluMetSerGlnMetLysProHisValTyrTrpLeuProVal	739
Db	1151	ARGGCTGTTCAACCCCAAGAGATGAGTCAGATGAAGGCCCATGTGTATGATCTCCCAAGTA	1210
Qy	740	GluGlyAsnGluProProAlaSerTrpHisLysTrpTrpAsnAlaLeuProProThrVal	759
Db	1211	GAAGGCATGAGCCCCCAGCTCTTTTCATTAAGACACAAACGACACTTCCCCCTACTGTT	1270
Qy	760	TrpLysTrpAsnProMetSerThrProLysTrpSerAsnProLysAla	775
Db	1271	TTCAAGATATTCAAATGCTCTACACCTAAGACTCAATCCCAAGCA	1318

RESULT 4

```

RES0001 4
US-10-125-470-4
; Sequence 4, Application US/10125470
; Publication No. US20020165377A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: ADAM Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PTO06P1
; CURRENT APPLICATION NUMBER: US/10/125,470
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US/09/712,907A
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: 60/136,388
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: 60/142,930
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 60/178,717
; PRIOR FILING DATE: 2000-01-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4

```

Db 851 ATTGAGAAAGCCATACAAATCAACCAATTCCTATCTAAGTGCAGAAAGACATGCTGTGTGT 910
QY 640 TrpHisLeuGlnTrpGlnTrpGluGlyTrpIleProProTrpTrpTrpSer 659
Db 911 GACCATGAGCTCCAGTGTCAATGTGAGAAAGGATGGATCCCTCCCGACTCGGATGACTCC 970
QY 660 SerValValTrpHisTrpSerIleValValGlyValLeuTrpProMetAlaValIleTrp 679
Db 971 TCAGTGGTCTTCACCTCTCCATTTGGTGTGGGTGCTGTTCCTCCATGGCGGTCACTTTT 1030
QY 680 ValValValAlaMetValIleArgHisGlnSerSerArgGluLysGlnLysTrpGln 699
Db 1031 STGGTGGTGTCTATGTGTAAATCCGGCCACACAGAGCTCCAGAGAAAGCAGAAAGATCAG 1090
QY 700 ArgProLeuSerThrThrGlyThrArgProHisLysGlnLysArgLysProGlnMetVal 719
Db 1091 AGCCCACTATCTACCACTGACACAGGCCACACAAACAGAGAGAAACCCGATGGTA 1150
QY 720 LysAlaValGlnProGlnGluMetSerGlnMetLysProHisValTrpLeuProVal 739
Db 1151 AAGGCTGTTCAACCCCAAGAGATGAGTCAGATGAAGCCCATGTGTATGATCTGCCAGTA 1210
QY 740 GluGlyAsnGlnProProAlaSerTrpHisLysTrpThrAsnAlaLeuProProTrpVal 759
Db 1211 GAAGGCAATGAGCCCCCAGCCTCTTTTCATAAAGACACAAACGCCTTCCCCCTACTGTT 1270
QY 760 TrpLysTrpAsnProXetSerThrProLysTrpSerAsnProLysAla 775
Db 1271 TTCAAGGATAATCAATGCTACACCTAAGGACTCAATCCAAAGCA 1318

RESULT 5

US-10-125-452-4
; Sequence 4, Application US/0125452
; Publication No. US20020173640A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: ADAM Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PTO06P2
; CURRENT APPLICATION NUMBER: US/10/125,452
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: 09/955,504
; PRIOR FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: 09/712,907
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: PCT/US00/14308
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/178,717
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/142,930
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 60/136,388
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 2047
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-125-452-4

Alignment Scores:

Pred. No.:	1,24e-135	Length:	2047
Score:	16.0/0.00	Matches:	345
Percent Similarity:	83.72%	Conservative:	20
Best Local Similarity:	79.13%	Mismatches:	71
Query Match:	34.17%	Indels:	0
DB:	14	Gaps:	0

US-10-726-148A-15 (1-775) x US-10-125-452-4 (1-2047)

QY 340 GluMetClyHisAsnTrpGlyMetTrpHisTrpTrpTrpSerTrpLysTrpProSerThr 359
Db 11 GAAATGGGCCCAACATTTTGAATTTTCAATGAGAGACTATTCTTGCAGAGTGTCTTCTTACA 70

QY 360 IleTrpValMetTrpLysAlaLeuSerTrpTrpIleProTrpTrpTrpSerTrpSer 379
Db 71 ATATGTGTGTATGAGCAAGACACTGAGCTTCTATATATACCCACAGACTTCAGTTCCTGAGC 130
QY 380 ArgLeuSerTrpTrpLysTrpTrpGluTrpLysLeuSerAsnTrpLeuTrpAsnAlaPro 399
Db 131 CGTCTCAGCTATGACAAGTTTTTTGAAGATAAATATCAAAATGCTCTTTTAATGCTCCA 190
QY 400 LeuProTrpTrpIleLeuSerThrProIleTrpGlyAsnGlnLeuValGluMetGlyGlu 419
Db 191 TTCCCTACAGATATCATATCCACTCCCAATTTGTGGGAACAGTTGGTGGAAATGGGAGAG 250
QY 420 TrpTrpTrpTrpGlyThrSerGluGluTrpThrAsnIleTrpTrpTrpAlaLysTrpTrp 439
Db 251 GACTGTGATTTGGGACATCTGAGGATGTACCAATATTTGCTGTGATGCTAAGACATGT 310
QY 440 LysIleLysAlaThrTrpGlnTrpAlaLeuGlyGluTrpTrpGluLysTrpGlnTrpLys 459
Db 311 AAAATCAAGCAACTTTTCAATGTGCATTAGGAGATGTTGTGAAAAATGCCAATTTAAA 370
QY 460 LysAlaGlyMetValTrpArgProAlaLysTrpGluTrpTrpLeuProGluMetTrpAsn 479
Db 371 AAGGCTGGGATGTGTGCAGCCAGCAAAAGATGAGTGCAGCCTGCCTGAAATGTGTAAT 430
QY 480 GlyLysSerGlyAsnTrpProTrpTrpArgTrpGlnValAsnGlyTrpProTrpHisHis 499
Db 431 GGTAAATCTGTAATTTCTCTGATGATAGATCCCAAGTCAATGCTTCCCTTGCATCAC 490
QY 500 GlyLysGlyHisTrpLeuMetGlyThrTrpProTrpLeuGlnGluGlnTrpTrpGluLeu 519
Db 491 GGGAAAGGCCACTGCTTGTATGGGACATGCCCCACACTGCGAGAGCAGTGCACAGAGCTG 550
QY 520 TrpGlyProGlyThrGluValAlaTrpLysSerTrpTrpAsnArgAsnGluGlyLysSer 539
Db 551 TGGGACCCAGAACTGAGGTTCAGATAAGTCATGTTTACCAACAGAAATGAAGTGGGTCA 610
QY 540 LysTrpGlyTrpArgArgValTrpTrpThrLeuIleProTrpLysAlaAsnTrpThr 559
Db 611 AAGTACGGGTACTTCCGACAGTGGATGACACACTCATTCCTCCAAAGCAATGATACC 670
QY 560 MetTrpGlyLysLeuTrpTrpGlnGlySerTrpAsnLeuProTrpLysGlyArgIle 579
Db 671 ATGTGTGGGAAGTGTCTCTCAAGGTGGGTGCGATAATTTGCCCTGCGAAAGGACGGATA 730
QY 580 ValTrpTrpLeuThrTrpLysThrTrpTrpProGluTrpThrSerGlnGluIleGlyMet 599
Db 731 GTGACTTCTCTGACATGTAAACATTTGATCCTTGACACACAGTCAAGAAATAGGCATG 790
QY 600 ValAlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpIleAsnAlaGluTrpValTrp 619
Db 791 GTGGCCAATGGAACTAAGTGTGGCGATAACAAGGTTTGCATTAATGCAAGATGTGCGAT 850
QY 620 IleGluLysAlaTrpLysSerThrAsnTrpSerSerLysTrpLysGlyHisAlaValTrp 639
Db 851 ATTGAGAAAGCCTACAAATCAACCAATTCCTCATCTAAGTGCAGAAAGACATGCTGTGT 910
QY 640 TrpHisGluLeuGlnTrpGlnTrpGluGlyTrpIleProProTrpTrpTrpTrpSer 659
Db 911 GACCATGAGCTCCAGTGTCAATGTGAGGAAGGATGGATCCCTCCCGACTCGGATGACTCC 970
QY 660 SerValValTrpHisTrpSerIleValValGlyValLeuTrpProMetAlaValIleTrp 679
Db 971 TCAGTGGTCTTCCACTTCTCATTGTGTGGGTGCTGTTCCTCCATGGCGGTCACTTTT 1030
QY 680 ValValValAlaMetValIleArgHisGlnSerSerArgGluLysGlnLysTrpGln 699
Db 1031 GTGGTGGTGTGTATGTTAATCCGGCACACAGAGCTCCAGAGAAAGCAGAAAGATCAG 1090
QY 700 ArgProLeuSerThrThrGlyThrArgProHisLysGlnLysArgLysProGlnMetVal 719
Db 1091 AGGCCACTATCTACCACTGGCAACAGGCCACACAAACAGAGAGAAACCCGATGGTA 1150


```
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 241
; LENGTH: 1652
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-299-241

Alignment Scores:
Pred. No.: 1.99e-67 Length: 1652
Score: 854.00 Matches: 179
Percent Similarity: 86.64% Conservativeness: 9
Best Local Similarity: 82.49% Mismatches: 29
Query Match: 18.12% Indels: 0
DB: 9 Gaps: 0

US-10-726-148A-15 (1-775) x US-09-925-299-241 (1-1652)
QY 559 ThrMetTrpGlyLysLeuTrpTrpGlnGlySerTrpAsnLeuProTrpLysGlyArg 578
DB 251 ACCATGTGGGAAGTTGTTCTGTCAAGTGGGTGGGATAATTGCTCCCTGGAAAGGACGG 310
QY 579 IleValThrTrpLeuThrTrpLysThrTrpProGluTrpThrSerGlnGluLeuGly 598
DB 311 ATAGTGACTTTCCTGACATGTARAACTTTCATCTGAGACACACAGTCAAGAAATARGC 370
QY 599 MetValAlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpIleAsnAlaGluTrpVal 618
DB 371 ATGTGGCCCAATGGAACCTAAGTGGCGATACCAAGTTTCATCAATGCAAGATGTGTG 430
QY 619 TrpIleGluLysAlaTyLysSerThrAsnTrpSerSerLysTrpLysGlyHisAlaVal 638
DB 431 GATATTGGAAGAGCTTACAAATCAACCAATTTGCTCATCTAAGTGCAGAGGACATCTGTG 490
QY 639 TrpTrpHisGluLeuGlnTrpGlnTrpGluGluGlyTrpIleProProTrpTrpTrp 658
DB 491 TGTGACCATGAGTCCAGTGTCAATGTGAGGAGGATGGATCCCTCCCGACTGGCATGAY 550
QY 659 SerSerValValTrpHisTrpSerIleValValGlyValLeuTrpProMetAlaValle 678
DB 551 TCCTCAGTGGTCTTCCACTTCTCCATTTGTTGGGTGGTCTCCCAATGGCGGTCAAT 610
QY 679 TrpValValAlaMetValIleArgHisGlnSerSerArgGluLysGlnLysLysTrp 698
DB 611 TTTGTGGTGGTGTATGTTATCGGCACAGGCTCCAGAGAACAGAGAAAGAT 670
QY 699 GlnArgProLeuSerThrThrGlyThrArgProHisLysGlnLysArgLysProGlnMet 718
DB 671 CAGAGGCCAYTATCTACCCCTGGCACCAGGCCACACAAACAGAGAAACCCAGATG 730
QY 719 ValLysAlaValGlnProGlnGluMetSerGlnMetLysProHisValTyTrpLeuPro 738
DB 731 GTAAAGCGTGTTCACCCCAAGAGATGAGTCAAGAGCCCATGTGTATGATCTGCCA 790
QY 739 ValGluGlyAsnGluProProAlaSerTrpHisLysTrpThrAsnAlaLeuProProThr 758
DB 791 GTAGAGGCAATGAGCCCCCAGCTCTTTTCATAAAGACACAAACGCACTTCCCCCTACT 850
QY 759 ValTrpLysTrpAsnProMetSerThrProLysTrpSerAsnProLysAla 775
DB 851 GTTTTCAAGGATAATCCAATGCTACACCTAAGGACTCAATTCACAAAGCA 901

RESULT 8
US-09-925-299-241
; Sequence 241, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
```

```
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 241
; LENGTH: 1652
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-299-241

Alignment Scores:
Pred. No.: 1.99e-67 Length: 1652
Score: 854.00 Matches: 179
Percent Similarity: 86.64% Conservativeness: 9
Best Local Similarity: 82.49% Mismatches: 29
Query Match: 18.12% Indels: 0
DB: 10 Gaps: 0

US-10-726-148A-15 (1-775) x US-09-925-299-241 (1-1652)
QY 559 ThrMetTrpGlyLysLeuTrpTrpGlnGlySerTrpAsnLeuProTrpLysGlyArg 578
DB 251 ACCATGTGGGAAGTTGTTCTGTCAAGTGGGTGGGATAATTGCTCCCTGGAAAGGACGG 310
QY 579 IleValThrTrpLeuThrTrpLysThrTrpProGluTrpThrSerGlnGluLeuGly 598
DB 311 ATAGTGACTTTCCTGACATGTARAACTTTCATCTGAGACACACAGTCAAGAAATARGC 370
QY 599 MetValAlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpIleAsnAlaGluTrpVal 618
DB 371 ATGTGGCCCAATGGAACCTAAGTGGCGATACCAAGTTTCATCAATGCAAGATGTGTG 430
QY 619 TrpIleGluLysAlaTyLysSerThrAsnTrpSerSerLysTrpLysGlyHisAlaVal 638
DB 431 GATATTGGAAGAGCTTACAAATCAACCAATTTGCTCATCTAAGTGCAGAGGACATCTGTG 490
QY 639 TrpTrpHisGluLeuGlnTrpGlnTrpGluGluGlyTrpIleProProTrpTrpTrp 658
DB 491 TGTGACCATGAGTCCAGTGTCAATGTGAGGAGGATGGATCCCTCCCGACTGGCATGAY 550
QY 659 SerSerValValTrpHisTrpSerIleValValGlyValLeuTrpProMetAlaValle 678
DB 551 TCCTCAGTGGTCTTCCACTTCTCCATTTGTTGGGTGGTCTCCCAATGGCGGTCAAT 610
QY 679 TrpValValAlaMetValIleArgHisGlnSerSerArgGluLysGlnLysLysTrp 698
DB 611 TTTGTGGTGGTGTATGTTATCGGCACAGGCTCCAGAGAACAGAGAAAGAT 670
QY 699 GlnArgProLeuSerThrThrGlyThrArgProHisLysGlnLysArgLysProGlnMet 718
DB 671 CAGAGGCCAYTATCTACCCCTGGCACCAGGCCACACAAACAGAGAAACCCAGATG 730
QY 719 ValLysAlaValGlnProGlnGluMetSerGlnMetLysProHisValTyTrpLeuPro 738
DB 731 GTAAAGCGTGTTCACCCCAAGAGATGAGTCAAGAGCCCATGTGTATGATCTGCCA 790
QY 739 ValGluGlyAsnGluProProAlaSerTrpHisLysTrpThrAsnAlaLeuProProThr 758
DB 791 GTAGAGGCAATGAGCCCCCAGCTCTTTTCATAAAGACACAAACGCACTTCCCCCTACT 850
QY 759 ValTrpLysTrpAsnProMetSerThrProLysTrpSerAsnProLysAla 775
DB 851 GTTTTCAAGGATAATCCAATGCTACACCTAAGGACTCAATTCACAAAGCA 901

RESULT 9
US-09-925-504-5
; Sequence 5, Application US/09955504
; Publication No. US20020182702A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
```

;; TITLE OF INVENTION: ADAM Polynucleotides, Polypeptides, and Antibodies

;; FILE REFERENCE: PTO06P2
;; CURRENT APPLICATION NUMBER: US/09/955,504
;; PRIOR FILING DATE: 2001-09-19
;; PRIOR APPLICATION NUMBER: 60/234,222
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: 09/712,907
;; PRIOR FILING DATE: 2000-11-16
;; PRIOR APPLICATION NUMBER: PCT/US00/14308
;; PRIOR FILING DATE: 2000-05-25
;; PRIOR APPLICATION NUMBER: 60/178,717
;; PRIOR FILING DATE: 2000-01-28
;; PRIOR APPLICATION NUMBER: 60/142,930
;; PRIOR FILING DATE: 1999-07-09
;; PRIOR APPLICATION NUMBER: 60/136,388
;; PRIOR FILING DATE: 1999-05-27
;; NUMBER OF SEQ ID NOS: 38
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 5
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-955-504-5

Alignment Scores:
Pred. No.: 2,45e-64 Length: 1704
Score: 820.00 Matches: 174
Percent Similarity: 84.33% Conservative: 9
Best Local Similarity: 80.18% Mismatches: 26
Query Match: 17.40% Indels: 8
DB: 9 Gaps: 1

US-10-726-148a-15 (1-775) x US-09-955-504-5 (1-1704)

QY 559 ThrMetTrpGlyLysLeuTrpTrpGlnGlyGlySerTrpAsnLeuProTrpLysGlyArg 578
Db 259 ACCATGTGGGAAAGTTCTTCTGTCAGGTGGGTGGGATAATTTGCCCTGGAAAGGACGG 318
QY 579 IleValThrTrpLeuTrpLysThrTrpTrpProGluTrpThrSerGlnGluLeGly 598
Db 319 ATAGTGACTTTCTGACATGTAAACAATTTGATCTCGAAGACACCAAGTCAAGAAATAGGC 378
QY 599 MetValAlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpIleAsnAlaGluTrpVal 618
Db 379 ATGGTGGCCATGGAACTAAGTGGCGATAACAAGGTTTGCATTAAATCCAGAAATGTGTG 438
QY 619 TrpIleGluLysAlaTyrlsSerThrAsnTrpSerLysTrpLysGlyHisAlaVal 638
Db 439 GATATTGAGAAAGCCTACAAATCAACCAATTTGATCTTAAGTGCAGAAAGGACATGCTGTG 498
QY 639 TrpTrpHisGluLeuGlnTrpGlnTrpGluGlyTrpIleProProTrpTrpTrp 658
Db 499 TGTGACCAAGAGCTCCAGTGAAGTGGAGGAGATGGATCCCTCCGACCTGCGATGAC 558
QY 659 SerSerValValTrpHisTrpSerIleValValGlyValLeuTrpProMetAlaValle 678
Db 559 TCCTCAGTGGTCTTCCACTTCTCCATTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 618
QY 679 TrpValValAlaMetValIleArgHisGlnSerSerArgGluLysGlnLysLysTrp 698
Db 619 TTTGG 678
QY 719 ValLysAlaValGlnProGlnGluMetSerGlnMetLysProHisValTrpTrpLeuPro 738
Db 739 -----ATGAGTCAGATGAAGCCCACTGTGATCATCTGCCA 774
QY 739 ValGluGlyAsnGluProProAlaSerTrpHisLysTrpThrAsnAlaLeuProProThr 758
Db 775 GTAGAAGGCATATGAGCCCCCAGCCTCTTTTCATAAAGACACAAACGCACTTCCCTCCCTACT 834

QY 759 ValTrpLysTrpAsnProMetSerThrProLysTrpSerAsnProLysAla 775
Db 835 GTTTTCAGGATAATCCATGTCTACACCTAAGGACTCAATTCATAAAGCA 885

RESULT 10

US-10-125-470-5
;; Sequence 5, Application US/10125470
;; Publication No. US20020165377A1
;; GENERAL INFORMATION:
;; APPLICANT: Ruben et al.
;; TITLE OF INVENTION: ADAM Polynucleotides, Polypeptides, and Antibodies
;; FILE REFERENCE: PTO06P1
;; CURRENT APPLICATION NUMBER: US/10/125,470
;; CURRENT FILING DATE: 2002-04-19
;; PRIOR APPLICATION NUMBER: US/09/712,907A
;; PRIOR FILING DATE: 2001-03-07
;; PRIOR APPLICATION NUMBER: 60/136,388
;; PRIOR FILING DATE: 1999-05-27
;; PRIOR APPLICATION NUMBER: 60/142,930
;; PRIOR FILING DATE: 1999-07-09
;; PRIOR APPLICATION NUMBER: 60/178,717
;; PRIOR FILING DATE: 2000-01-28
;; NUMBER OF SEQ ID NOS: 24
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 5
;; LENGTH: 1704
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-125-470-5

Alignment Scores:
Pred. No.: 2,45e-64 Length: 1704
Score: 820.00 Matches: 174
Percent Similarity: 84.33% Conservative: 9
Best Local Similarity: 80.18% Mismatches: 26
Query Match: 17.40% Indels: 8
DB: 14 Gaps: 1

US-10-726-148a-15 (1-775) x US-10-125-470-5 (1-1704)

QY 559 ThrMetTrpGlyLysLeuTrpTrpGlnGlyGlySerTrpAsnLeuProTrpLysGlyArg 578
Db 259 ACCATGTGGGAAAGTTCTTCTGTCAGGTGGGTGGGATAATTTGCCCTGGAAAGGACGG 318
QY 579 IleValThrTrpLeuTrpLysThrTrpTrpProGluTrpThrSerGlnGluLeGly 598
Db 319 ATAGTGACTTTCTGACATGTAAACAATTTGATCTCGAAGACACCAAGTCAAGAAATAGGC 378
QY 599 MetValAlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpIleAsnAlaGluTrpVal 618
Db 379 ATGGTGGCCATGGAACTAAGTGGCGATAACAAGGTTTGCATTAAATCCAGAAATGTGTG 438
QY 619 TrpIleGluLysAlaTyrlsSerThrAsnTrpSerLysTrpLysGlyHisAlaVal 638
Db 439 GATATTGAGAAAGCCTACAAATCAACCAATTTGATCTTAAGTGCAGAAAGGACATGCTGTG 498
QY 639 TrpTrpHisGluLeuGlnTrpGlnTrpGluGlyTrpIleProProTrpTrpTrp 658
Db 499 TGTGACCAAGAGCTCCAGTGAAGTGGAGGAGATGGATCCCTCCGACCTGCGATGAC 558
QY 659 SerSerValValTrpHisTrpSerIleValValGlyValLeuTrpProMetAlaValle 678
Db 559 TCCTCAGTGGTCTTCCACTTCTCCATTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 618
QY 679 TrpValValAlaMetValIleArgHisGlnSerSerArgGluLysGlnLysLysTrp 698
Db 619 TTTGG 678
QY 699 GlnArgProLeuSerThrThrGlyThrArgProHisLysGlnLysArgLysProGlnMet 718
Db 679 CAGAGGCCCATATGAGCCCCCAGCCTCTTTTCATAAAGACACAAACGCACTTCCCTCCCTACT 738

QY 57 LeuLysTyrIshYSMetThrLeuAsnGlyLysIleAlaValLeuTyrLeuLysAsnLys 76
D5 :
257 ATGCATATGAATTAAAGATGAACGGAGAGCCAGTCGTCTTCCTCACCTGGAAAAATAAAA 316

QY 77 AsnLeuLeuAlaProGlyTyrThrGluThrTyrTyrAsnSerThrGlyLysGluIleThr 96
D5 :
317 AGACTTTTTCMAAAGAATTACCGCAGACTCATTTCCOCTGATGCCAGACAATTCACA 376

QY 97 ThrSerProGlnIleMetTrpTrpTyrTyrGlnGlyHisIleLeuAsnGlysVal 116
D5 :
377 ACATACCCCATGATTGAGGATCACTGCTATTATCATGGCCCATCCAGATGATGCTGAC 436

QY 117 SerTrpAlaSerIleSerThrTrpArgGlyLeuArgGlyTyrTrpSerGlnGlyTrpGln 136
D5 :
437 TCAACTGCAAGCATCACTGTCATGCACCGTTTGAAAGGACATTTCAAGCTTCAAGGGAG 496

QY 137 ArgTyrTrpIleGlnProLeuSerProIleHisArgTrpGlyGlnHisAlaLeuTrp 156
D5 :
497 ATGTACCCTCATTAAGACCTTTGAAGCTTCCCGAC-----AGTGAAGCCCATGCAGTCTAC 550

QY 157 LysTyrAsnProTrpGluLysAsnTyrTrpSerThrTrpGlyMetTrpGlyValLeuTrp 176
D5 :
551 AAATATGAAAACATAGNAAAAGGATGAGGCCCCCC---AAATGTGTGGGTAAACC--- 604

QY 177 AlaHisTrpLeuGlnInAsnIleAlaLeuProAlaThrLysLeuValLysLeuLysTrp 196
D5 :
605 CAGAAITGGGAATCATATGAG-----CCCATCAAAAAGGCCCTTTCAGTTAAATCTT 655

QY 197 ArgLysValGln-----GluHisGluLysTyrIleGluTyrTyrLeuValLeu 212
D5 :
656 ACTCCTGAAACAACAGCATCTTGGATGCCAAAATAACGTTGAGTTTGTGTAGTTCGT 715

QY 213 TrpAsnGlyGluTrpLysArgTyrAsnGluAsnGlnTrpGluIleArgLysArgValTrp 232
D5 :
716 GACCATGGATGTACACAAATACAAAGATGATTTAGATAGATATAAAACAGAAATATAT 775

QY 233 GluMetAlaAsnTyrValAsnMetLeuTyrLysLysLeuAsnThrHisValAlaLeuVal 252
D5 :
776 GAAATTCACACACTATGATGAGATTTACATACCTTTTGAATATTGCTGTCGCACTGGTT 835

QY 253 GlyMetGluIleTrpThrTrpLysTrpLysIleLysIleThrProAlaAlaSerTrpThr 272
D5 :
836 CACCTAGAAATTTGGTCCAACAGAGATTTGATTAATGTGTCAACAGCAGCGTGATCT 895

QY 273 LeuGluAsnTrpSerLysTrpArgGlySerValLeuSerArgArgLysArgHisTrpIle 292
D5 :
896 TTGGGCTCATTTGGAGATGAGAGACAGACAGATTTGCTGAGGCACHAAGCTCATGATAAT 955

QY 293 AlaGlnLeuIleThrAlaThrGluLeuAlaGlyThrThrValGlyLeuAlaTrpMetSer 312
D5 :
956 GCCTCAGTTACTCAGCACCACTGACTTCGATGGAGACACTGTAGSATTTGGCTTATATAAGC 1015

QY 313 ThrMetTrpSerPro----TyrSerValGlyValValGlnTrpHisSerTrpAsnLeuLeu 331
D5 :
1016 AGCATGTGCCAACCGCAGCTTCTGTAGGAGTTATTTCAGGAACATAGACAAACAAATCTC 1075

QY 332 ArgValAlaGlyThrMetAlaHisGluMetGlyHisAsnTrpGlyMetTrpHisTrpTrp 351
D5 :
1076 ITTAGTCAGTTCAATGGCCCATGAGATGGTCTCAATCTGGGCATGAGTCATGATGGA 1135

QY 352 TyrSerTrpLysTrpProSerThrIleTrpValMetTrpLysAlaLeuSerTrpTyrIle 371
D5 :
1136 AATCAGTGTCAATTGTGTGCTCCCTCGTCATTATGCTGAAAGACTAAGCCACCAACCT 1195

QY 372 ProThrTrpTrpSerSerTrpSerArgLeuSerTyrTrpLysTrpTrpGluTrpLysLeu 391
D5 :
1196 TCCACACAGTTTCAGCGATTGTAGTGAGGAATATTGTCCGACGATATCTTAAAAATCGTAGA 1255

QY 392 SerAsnTrpLeuTrpAsnAlaProLeuProThrTrpIleIleSerThrProIleTrpGly 411
D5 :
1256 CCACAAATGCATCTCAATGAACCTTCTCTGCACAGATATTGTCTTCACTCCAGTTGTGGA 1315

QY 412 AsnGlnLeuValGluMetGlyGluTrpTrpTrpGlyThrSerGluGluTrpThrAsn 431

[illegible]

Score: 755.00 Matches: 245
Percent Similarity: 48.11% Conservative: 136
Best Local Similarity: 30.93% Mismatches: 335
Query Match: 16.02% Indels: 76
DB: 13 Gaps: 20

US-10-726-148A-15 (1-775) x US-10-232-972B-3 (1-2634)

Qy 1 MetLeuGlnGlyLeuLeuProValSerLeuLeuSerValAlaValSerAlaIleLys 20
Db 85 ATGCTTCGGGCTGATATTCTGATGATTTTACTC-----ATTCTTCAGGTAAA 135
Qy 21 Glu-----LeuProGlyValLysLysTyrgluValValTyProIleAlaLeuHisPro 38
Db 136 GAAAGTTTCATCTTGGAGTAGAGGGTCAACAACTGGTTGCTCTTAAGAGCTCTCCTCTG 195
Qy 39 LeuHisLysArgGluAlaLysGluProGlnGlnGlu-----GlnTrpGluThr 55
Db 196 ATACAGAGGAGAGATACCTGGACACACCATGATGACATCTACTCAAAACGTATGAGAA 255
Qy 56 GluLeuLysTyrgluMetThrIleAsnGlyLysIleAlaValLeuTyrgluLysLysAsn 75
Db 256 GAATTTGTTGATGAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 315
Qy 76 LysAsnLeuLeuAlaProGlyTyrgluThrTyrgluThrTyrgluThrTyrgluThrTyrgluThr 95
Db 316 AGGAGTTCCTAGGCTCAATTAAGTGAACATCTACTCCATCAAGAGGAGGAGCGTTC 375
Qy 56 ThrThrSerProGlnIleMetTrpTrpTyrgluTyrgluGlnGlyHisIleLeuAsnGluLys 115
Db 376 ACCAGGATCCTCAGATCATGATCATGTTTATACCAAGGATCCATAGTACAGAAATAT 435
Qy 116 ValSerTrpAlaSerIleSerThrTrpArgGlyLeuArgGlyTyrgluThrSerGlnGlyTrp 135
Db 436 GATTACGCTGCCATATACGATGATGTAATAGTCTAAGGGGATCTCCAGAAATAAAGCAG 495
Qy 136 GluArgTyrgluLeuProLeuSerProIleHisArgTrpGlyGlnGlnHisAlaLeu 155
Db 496 CAAAGACCTCATTTGACCAAGTGAAG-----TACTCAGATGAGGGAGACATTTGGTG 549
Qy 156 TrpLysTyrglu-----ProTrpGluLysAsnTyrgluTrpSerThrTrpGlyMetTrp 172
Db 550 TTCAATATAAATCACTGAGGGTCCGCTATGCTGCTCAATTTATCTCTGTAACA----- 597
Qy 173 GlyValLeuTrpAlaHisTrpLeuGlnGlnAsnIleAlaLeuProAlaThrLysLeuVal 192
Db 598 -----GAGCTTAATTTTACCAGAAATACTGTTCCAGGGGATATGAAATCTGAAGAA 648
Qy 193 LysLeuLysTrpArgLysValGlnGlnHisGlnLysTyrgluTyrgluLeuValLeu 212
Db 649 GACTCCAAAATAAAGGCATCCATGAT--GAAAGATATGTTGAATGTTCTCATTTGCTCT 705
Qy 213 TrpAsnGlyGluTrpLysArgTyrgluAsnGlnTrpGluIleArgLysArgValTrp 232
Db 706 GATGATAGTGTATCGCAGAAATAGGTCATCCTCACAATAAATAAAGGAAACCGAATTTGG 765
Qy 233 GluMetAlaAsnTyrgluValSerMetLeuTyrgluLysLysLeuAsnThrHisValAlaLeuVal 252
Db 766 GGAATGTGCTAATTTTGCAACATGATTTATAAATAAATAAATAAATAAATAAATAAATAA 825
Qy 253 GlyMetGluIleTrpThrTrpLysTyrgluLysIleTrpProAsnAlaSerTrpThr 272
Db 826 GGCATTGAAATATGACACATGAAGATAAATAAATAAATAAATAAATAAATAAATAAATAA 885
Qy 273 LeuGlnAsnTrpSerLysTrpArgLysSerValLeuSerArgTyrgluHisTrpIle 292
Db 886 TTATGTGCTTTTTCATTTTGGCAAGAAAGATCTCTTAAACACGGAAGGATTTTGTATCAT 945
Qy 293 AlaGlnLeuIleThrAlaThrGluLeuAlaGlyThrThrValGlyLeuAlaTrpMetSer 312
Db 946 GTTGTATTTACTCAGTGGGAGTGGCTCTACTCAATGTCAGAGGAAATTTCTATCCAGGG 1005
Qy 313 ThrMetTrpSerPro---TyrSerValGlyValValGlnTrpHisSerTrpAsnLeuLeu 331

Db 1006 GGTAATGTCCTGCCCTATTATTCCACCAAGTATCATTAAG-----GATCTTTTA 1053
Qy 332 Arg-----ValAlaGlyThrMetAlaHisGluMetGlyHisAsnTrpGlyMet 347
Db 1054 CTGCACACAAACATAAATTGCAACACAGAAATGCACATCAACTGGGGCATAACCTTGGGATG 1113
Qy 348 TrpHisTrpTrpTyrgluSerTrpLysTrpProSerThrIleTrpValMetTrpLysAlaLeu 367
Db 1114 CAGCATGACAGAGTTTCCCATGACCTGCTCTTCAGGAAATCGTGATCGACAGTGAATGGA 1173
Qy 368 SerTrpTyrgluProThrTrp---TrpSerTrpSerArgLeuSerTyrgluTrpLysTrp 386
Db 1174 AGC-----ATCTGCACTGAATTCAGTAATATGCAGCAAAACCAATATCACCAAGTAC 1227
Qy 397 TrpGluTrpLysLeuSerAsnTrpLeuTrpAsnAlaProLeuProTrpThrTrpIleLeuSer 406
Db 1228 TTGAAGGATTATAAGCCAAACATGCTCAACATTCCTTCCATTCCATTCAATTTTCATGAT 1287
Qy 407 ThrProIleTrpGlyAsnGlnLeuValGluMetGlyGluTrpTrpTrpTrpGlyThrSer 426
Db 1288 TTCCAAATTTTGTGAAACAAAGAGTTGATGAGGGTGAAGAGTGTGACTGTGGCCCTGCT 1347
Qy 427 GluGluTrpThrAsnIleTrpTrpAlaLysThrTrpLysIleLysAlaThrTrpGln 446
Db 1348 CAGAGTGTACTAATCTGCTGTGATGCAACACATCTGTACTGAAGCCAGGATTTACT 1407
Qy 447 TrpAlaLeuGlyGluTrpTrpGluLysTrpGlnTrpLysLysAlaGlyMetValTrpArg 466
Db 1408 TGTGCAGAAAGAGAAATGCTGTGAATCTTGTGAGATAAATAAAGCAGGGTCTCATATGCA 1467
Qy 467 ProAlaLysTrpGluTrpTrpLeuProGluMetTrpAsnGlyLysSerGlyAsnTrpPro 486
Db 1468 CCGCGAAGAGATGATGATTTCTGAGATGTCACCTGGCCACTCGCTGCTGCTGCTGCT 1527
Qy 487 TrpArgTrpGlnValAsnGlyTrpProTrpHisGlyLysGlyHisTrpLeuXet 506
Db 1528 AAGGACCACTTCAGGTCATGATTTCTTGCAGAACTCAGAAAGGCTACTGTTTCTATG 1587
Qy 507 GlyThrTrpProThrLeuGlnGluGlnTrpThrGluLeuTrpGlyProGlyThrGluVal 526
Db 1588 GGGAAATGTCACACTCGTGGAGATCAGTCTCTGAACTATTGTATGATGATGATGATGAT 1647
Qy 527 AlaTrpLysSerTrpTyrgluAsnArgSerGluGlyGlySerLysTyrgluTyrgluArgArg 546
Db 1648 AGTCATGATATCTCTCAAGATCAATACAAAGAAATAAATTTGGATCTGCAAAAC 1707
Qy 547 ValTrpTrpThrLeuIleProTrpLysAlaAsnTrpMetTrpGlyLysLeuTrpTrp 566
Db 1708 AAGGAAACACAGATTTCTCCCTGTGAGGAGAAAGATGTCAGATGTGAAAGATCTACTGTC 1767
Qy 567 GlnGlyGly-----SerTrpAsnLeu-----Pro 574
Db 1768 ACTCGAGGGAGCTTCTCTCTCTCTGAGAGAACAGACTTATCATCTTCAAGGATCCC 1827
Qy 575 TrpLysGlyArgIleValThrTrpLeuThrTrpLysThrTrpTrpProGluTrpThrSer 594
Db 1828 CAGAGAAATGCTACTGTC-----AAATGCAAACTATTTTATATACCATGATCT 1878
Qy 595 GlnGluIleGlyMetValAlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpIleAsn 614
Db 1879 ACAGACATTTGCTGTGGTGGTGTGAGCAACAAATGTGAGAGGAAATGCTGTGCAACAAT 1938
Qy 615 AlaGluTrpValTrpIleGluLysAlaTyrgluSerThrAsnTrpSerSerLysTrpLys 634
Db 1939 GGTGAATGTCTAAACATGGAAGAGGTCTATATCTCAACCAATGCCCTCTCAGTGCAAT 1998
Qy 635 GlyHisAlaValTrpTrpHisGluLeuGlnTrpGlnTrpGluGluGlyTrpIleProPro 654
Db 1999 GAAATCTCTGGATGGCCACGGACTCCAGTCCACTGTGAGGAGGACAGGACCTGTA 2058
Qy 655 TrpTrpTrpTrpSerSerValValTrpHisTrpSerIleValValGlyValLeuTrpPro 674

```

Db      2059  GCCTGGAAGAAACCTTACATGTTACCAATATACACATCTTGGTGTCTTGTCTGCTTCCTG 2118
QY      675  MetAlaValIleTrrpValValIleMetValIleArgHisGlnSerSerArgGluLys 694
Db      2119  GTTATGTGCGTATCGGAGTCTTACTATTAGTTCGTACCGAAATGTATCAAGTTG 2178
QY      695  GlnLysLysTrrpGlnArgProLysSerThrThrGly----- 706
Db      2179  AAGCAAGTTCAGAGCCCACTACAGAAACCTGGGAGTGGAGAGCAACAAAGATACATTGGT 2238
QY      707  -----ThrArgPro-----HisLysGlnLysArgLys 715
Db      2239  GATGAGCAGCAGATAGAGTCTGAGCAATCTGCCAGAAATTCATTCTCTAAATAGAACT 2298
QY      716  ProGlnMetValLysAlaValGlnProGlnGluMetSerGlnMetLysProHisValTyr 735
Db      2299  CCAGAATCCTTGGAAAGCGTG---CCCACTAGTTTTTCA-----AGTCCCACTACATC 2349
QY      736  TrpLeuProValGluGlyAsnGluProProAlaSer 747
Db      2350  ACACTG-----AAACTGCAAGT 2367

RESULT 14
US-10-232-972B-1
; Sequence 1, Application US/10232972B
; Publication No. US20040043387A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Hwan-Wun
; APPLICANT: Sun, Guang-Huan
; APPLICANT: Lin, Yu-Chi
; APPLICANT: Chang, Sun-Yuan
; TITLE OF INVENTION: Nucleic Acid Molecules and Polypeptides Related to h-ADAM 7
; FILE REFERENCE: 87161178-562001
; CURRENT FILING DATE: US/10/232, 972B
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 3451
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: (85)..(2409)
US-10-232-972B-1

Alignment Scores:
Pred. No.:      4,45e-58      Length:      3451
Score:          755.00      Matches:      245
Percent Similarity: 48.11%      Conservative: 136
Best Local Similarity: 30.93%      Mismatches:  335
Query Match:     16.02%      Indels:       76
DB:              13          Gaps:         20

US-10-726-148a-15 (1-775) x US-10-232-972B-1 (1-3451)
QY      1  MetLeuGlnGlyLeuProValSerLeuLeuLeuSerValAlaValSerAlaIleLys 20
Db      85  ATGCTCCCGGTGTATATCTTGATGATTTACTC-----ATTCCTCAGGTAA 135
QY      21  Glu-----LeuProGlyValLysLysTyrGluValValTyrProIleArgLeuHisPro 38
Db      136  GAAAGATTCCTCTCGATGAGGTCAACACTGTTCTGTCCTTAAAGAGCTTCTCTG 195
QY      39  LeuHisLysArgGluAlaLysGluProGluGlnGlnGlu-----GlnTrrpGluThr 55
Db      196  ATACAGACGAGGATACTGCACACCCCATGATGATGACATACTGAAACGATGAAGAA 255
QY      56  GluLeuLysTyrLysMetThrIleAsnGlyLysIleAlaValLeuTyrLeuLysLysAsn 75
Db      256  GAATTTGTTATGAATAAATCAATAGAAAACCTTAGTCTCTTCATCTTCAATATCC 315
QY      76  LysAsnLeuLeuAlaProGlyTyrThrGluThrTyrTyrAsnSerThrGlyLysGluIle 95

```

```

Db      316  AGGAGTTCCTAGCTCAAAATACAGTGAACATCTACTCCATGAAGAGGAGCGTTC 375
QY      96  ThrThrSerProGlnIleMetTrrpTrrpTrrpTrrpTrrpTrrpTrrpTrrpTrrp 115
Db      376  ACCAGGCATCCTCAGATCATGATCATTTGTTTATACAGAGATCCATAGTACAGATAT 435
QY      116  ValSerTrrpAlaSerIleSerThrTrrpArgGlyLeuArgGlyTrrpTrrpSerGlnGlyTrrp 135
Db      436  GATTCAGCTGCAGATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 495
QY      136  GlnArgTrrpTrrpIleGluProLysSerProIleHisArgTrrpGlyGlnGlnHisAlaLeu 155
Db      496  CAAAGATACCTCATTCAGACCATGAAA-----TACTCAGATCAGGAGAGAACATTTGGTG 549
QY      156  TrrpLysTrrpAsn-----ProTrrpGluLysAsnTrrpTrrpSerThrTrrpGlyMetTrrp 172
Db      550  TTCAATATAAACCCTGAGGGTGCCTATGTTGTCCTATTTCTCTGTACA----- 597
QY      173  GlyValLeuTrrpAlaHisTrrpLeuGlnGlnAsnIleAlaLeuProAlaThrLysLeuVal 192
Db      598  -----GAGCTTAAATTTTACCAGAAAACCTGTTCCAGGGGATATGATCTGAAGAA 648
QY      193  LysLeuLysTrrpArgLysValGlnGlnHisGluLysTrrpIleGluTrrpTrrpLeuValLeu 212
Db      649  GACTCCAAAATAAAAGCATCCATGAT---GAAAGATATGTTGATTTGTTTCTTGTGCT 705
QY      213  TrrpAsnGlyGluTrrpLysArgTrrpAsnGluAsnGlnTrrpGluLysArgValTrrp 232
Db      706  GATGATACTGTGTATGCGAGAAATGTCATCTCACAATAAACTAAAGAACCGAATTTGG 765
QY      233  GluMetAlaAsnTrrpValAsnMetLeuTrrpLysLysLeuAsnThrHisValAlaLeuVal 252
Db      766  GGAATGTCATTTTGTCAACATGATTTATAAACCTTAAACATCCATGACGCTGGTT 825
QY      253  GlyMetGluIleTrrpTrrpTrrpLysLysIleLysIleThrProAsnAlaSerTrrpThr 272
Db      826  GGCATTGAAATATGGACACATGAAGATAAAATAGAACTATATCAAAATATAGAACTACC 885
QY      273  LeuGluAsnTrrpSerLysTrrpArgGlySerValLeuSerArgArgLysArgHisTrrpIle 292
Db      886  TTATTCGCGTTTTCATTTTGGCAAGAAAGATCCTTAAACACAGAGGATTTTGATCAT 945
QY      293  AlaGlnLeuIleThrAlaThrGluLeuAlaGlyThrThrValGlyLeuAlaTrrpMetSer 312
Db      946  GTTGTATTTACTCAGTGGAGTGGCTCTACTCATGTGCAAGGAAATTTCTTATCCAGGG 1005
QY      313  ThrMetTrrpSerPro---TyrSerValGlyValValGlnTrrpHisSerTrrpAsnLeuLeu 331
Db      1006  GGTATGTGCTGCGCTTATTATTCACCATGATCATTAAG-----GATCTTTTA 1053
QY      332  Arg-----ValAlaGlyThrMetAlaHisGluMetGlyHisAsnTrrpGlyMet 347
Db      1054  CCTGACACAAACATAATTCGCAACAGATGCGCATCACTCACTGGGGCATCAACCTTGGGATG 1113
QY      348  TrrpHisTrrpTrrpSerTrrpLysTrrpProSerThrIleTrrpValMetTrrpLysAlaLeu 367
Db      1114  CAGCATGACGAGTTCCTCATGCACTGCTCTCTCAGGAAAATGCTGATGACAGTATGGA 1173
QY      368  SerTrrpTrrpIleProThrTrrp---TrrpSerSerTrrpSerArgLeuSerTrrpTrrpLysTrrp 386
Db      1174  AGC-----ATTCCTGCACTGAAATTCAGTAAATGCGCCAAACCAACCAATACCACTAGTAC 1227
QY      387  TrrpGluTrrpLysLeuSerAsnTrrpLeuTrrpAsnAlaProLeuProThrTrrpIleIleSer 406
Db      1228  TTGAAGAGATTATAAGCCAAACATGCACTCAACATTCCTTCTTACAAATTTTCATGAT 1287
QY      407  ThrProIleTrrpGlyAsnGlnLeuValGluMetGlyGluTrrpTrrpTrrpTrrpTrrpTrrp 426
Db      1288  TTCCAAATTTTGTGGAACCAAGAGTTGGATGAGGAGTGAAGAGTGTGACTGTGCGCTGCT 1347
QY      427  GluGluTrrpThrAsnIleTrrpTrrpTrrpAlaLysThrTrrpLysIleLysAlaThrTrrpGln 446

```

Db 1348 CAGAGTGACTAACTCTCTGTGATGCACACACATGTGTACTGAGCCAGGATTACT 1407
Qy 447 TrpAlaLeuGlyGlnTrpGluLysTrpGlnTrpLysLysAlaGlyMetValTrpArg 466
Db 1408 TGTGCAGAGGAGTGTGTGATCTTGTTCAGATAAAAAAGCAGGTCCTCATGCAGA 1467
Qy 467 ProAlaLysTrpGlnTrpLeuProGluMetTrpAsnGlyLysSerGlyAsnTrpPro 486
Db 1468 CCGCGAAAGATGAATGTGATTTTCCTGAGATGTCACCTGCCCTGCCCTGCTCT 1527
Qy 487 TrpTrpArgTrpGlnValAsnGlyTrpProTrpHisGlyLysGlyHisTrpLeuMet 506
Db 1528 AAGACAGTTCAGGTCATGATTCCTTCAGAACTCAGAAAGCTACTGTTTCATG 1587
Qy 507 GlyThrTrpProThrLeuGlnGluTrpTrpGluLeuTrpGlyProGlyThrGluVal 526
Db 1588 GGGAAATGTCACACTCGTGGATCAGTGTCTGAACTATTGTATGATGATGCAATAGAG 1647
Qy 527 AlaTrpLysSerTrpTrpAsnArgAsnGluGlyGlySerLysTrpGlyTrpArgArg 546
Db 1648 AGTCATGATCTCTCAAGATGAATACAAAGGAATAAATTTGGATACCTGCAAAAC 1707
Qy 547 ValTrpTrpThrLeuLeuProTrpLysAlaAsnTrpThrMetTrpGlyLysLeuTrpTrp 566
Db 1708 AAGGAAACAGATTCTCTCTCTGTGAGAGAAAGATGTCAGATGTGGAAATCTACTGC 1767
Qy 567 GlnGlyGly-----SerTrpAsnLeu-----Pro 574
Db 1768 ACTGGAGGGAGCTTCT 1827
Qy 575 TrpLysGlyArgLeuValThrTrpLeuTrpThrTrpLysThrTrpTrpProGluTrpThrSer 594
Db 1828 CAGAAGATGCTACTGTC-----AAATGCAAACTACTTTTATACCATGATTC 1878
Qy 595 GlnGluLeuGlyMetValAlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpIleAsn 614
Db 1879 ACAGACATTCGCTGTGTGGTGTGAGGAAATGTGAGAGGAATGTGTGCAACAT 1938
Qy 615 AlaGluTrpValTrpLysGlyLysSerThrLysSerThrAsnTrpSerSerLysTrpLys 634
Db 1939 GTTGAATGTCTAAACATGAGAAAGTCTATATCTCAACCAATTCCTCTCTCAAGTCAAT 1998
Qy 635 GlyHisAlaValTrpTrpHisGluLeuGluTrpGlnTrpGluGluGlyTrpIleProPro 654
Db 1999 GAAATCTGTGGATGCGGAGCTCCAGTCCACCTGTGAGGAGGACAGGACCTGTA 2058
Qy 655 TrpTrpTrpTrpSerSerValValTrpHisTrpSerIleValValGlyValLeuTrpPro 674
Db 2059 GCCTGTGAAGAACTTACATGTTACCAATATACCACTTGTGTGTGTGTGTGTGTGT 2118
Qy 675 MetAlaValIleTrpValValAlaMetValIleArgHisGlnSerSerArgGluLys 694
Db 2119 GTTATTGTGCGTATCGAGTCTTATATCTATAGTTAGTTCTGTTACCGAAAAATGTATCAAGTTG 2178
Qy 695 GlnLysLysTrpGlnArgProLeuSerThrThrGly----- 706
Db 2179 AAGCAAGTTGAGCCACCTACAGAAACCTGGAGTGGAGAACAAAGATCTTTGGT 2238
Qy 707 -----ThrArgPro-----HisLysGlnLysArgLys 715
Db 2239 GATGAGCAGCAGATAAGGACTGAGCCCAATCTGCCAGAAATTCATTTCTCAATAGAACT 2298
Qy 716 ProGlnMetValLysAlaValGlnProGlnGluMetSerGlnMetLysProHisValTrp 735
Db 2299 CCAGAACTCTTGGAAAGCCCTG---CCCACTAGTTTTCAT-----AGTCCCCCACTACATC 2349
Qy 736 TrpLeuProValGluGlyAsnGluProProAlaSer 747
Db 2350 AACTG-----AAACCTGCAGT 2367

RESULT 15

US-09-908-193-9

; Sequence 9, Application US/09908193

Publication No. US20020192748A1
GENERAL INFORMATION:
APPLICANT: RASTELLI, LUCA
APPLICANT: SHIMKETS, RICHARD A.
APPLICANT: ZERHUSEN, BRYAN
APPLICANT: MALYANKAR, URIEL M.
APPLICANT: PADIGARU, MURALIDHARA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
FILE REFERENCE: 21402-062
CURRENT APPLICATION NUMBER: US/09/908,193
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: 60/220,273
PRIOR FILING DATE: 2000-07-24
PRIOR APPLICATION NUMBER: 60/221,650
PRIOR FILING DATE: 2000-07-28
PRIOR APPLICATION NUMBER: 60/221,233
PRIOR FILING DATE: 2000-07-27
PRIOR APPLICATION NUMBER: 60/220,912
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/218,875
PRIOR FILING DATE: 2000-07-18
PRIOR APPLICATION NUMBER: 60/218,870
PRIOR FILING DATE: 2000-07-18
PRIOR APPLICATION NUMBER: 60/218,901
NUMBER OF SEQ ID NOS: 74
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 9
LENGTH: 2434
TYPE: DNA
ORGANISM: Unknown Organism
FEATURE:
OTHER INFORMATION: Description of Unknown Organism: NOVX nucleic acid
OTHER INFORMATION: with homology to ADAM-like protein
US-09-908-193-9

Alignment Scores:

Pred. No.: 7,32e-58 Length: 2434
Score: 750.50 Matches: 251
Percent Similarity: 47.59% Conservative: 134
Best Local Similarity: 31.03% Mismatches: 347
Query Match: 15.93% Indels: 77
Dbs: 9 Gaps: 21

US-10-726-148A-15 (1-775) x US-09-908-193-9 (1-2434)

Qy 1 MetLeuGlnGlyLeuLeuProValSerLeuLeuSerValAlaValSerAlaIleLys 20
Db 51 ATGCTTCCGGGTGATATTCTTGTATGATTACTC-----ATTCTTCAGGTAAA 101
Qy 21 Glu-----LeuProGlyValLysLysTrpGluValValTrpProIleArgLeuHisPro 38
Db 102 GAAAGTTTCATCTCTGGAGTAGAGGGTCAACACAGTGTCTCTCTCTCTCTCTCTCT 161
Qy 39 LeuHisLysArgGluAlaLysGluProGluGlnGlnGlu-----GlnTrpGluThrGlu 56
Db 162 ATACGAGAGGAGATACCTGGACACACCATCATGATGACATATAAAACGTATGAGAGAA 221
Qy 57 LeuLysTrpLysMetThrIleAsnGlyLysIleAlaValLeuTyf---LeuLysLysAsn 75
Db 222 TTGTTGTATGAAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 281
Qy 76 LysAsnLeuLeuAlaProGlyTrpThrGluThrTrpTrpAsnSerThrGlyLysGluIle 95
Db 282 AGGAGGTTCCTAGCTCAAAATTACAGTGAACATCTTACTCTCATCTCTCTCTCTCTCT 341
Qy 96 ThrThrSerProGlnIleMetTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrp 115
Db 342 ACCAGGATCTCTAGATCATGGAACATCTGTACTCTATAAGGAACATCTCTATAAGAAAG 401
Qy 116 ValSerTrpAlaSerIleSerThrTrpArgGlyLeu---ArgGlyTrpTrpSerGlnGly 134
Db 402 AATTCTGTGCCAGCATCAGTACTTGTGACGGGTTGAGGAGGGGATCTTCTCAGATAAAC 461

Qy 135 TrpGlnArgTyrTrpIleGluProLeuSerProIleHisArgTrpGlyGlnGluHisAla 154
Db 462 GACCAGAAATACCTCATTTGAACAGTGAA-----TACTCAGATGAGGAGAACATTG 515
Qy 155 LeuTrpLysTyrAsn-----ProTrpGluLysAsnTyrTrpSerThrTrpGlyMet 171
Db 516 GTGTTCAAATAAACCCTGAGGCTGCGGTATGCGCAATTATTCCTGTACA----- 566
Qy 172 TrpGlyValLeuTrpAlaHisTrpLeuGlnGlnAsnIleAlaLeuProAlaThrLysLeu 191
Db 567 -----GAGCTTAATTTTACCAGAAAANAAGTGTCCAGGGGATAATGAATCTGAA 614
Qy 192 ValLysLeuLysTrpArgLysValGlnGluHisGluLysTyrIleGluTyrTrpLeuVal 211
Db 515 GAAGACTCCAAAATAAACAACAGGCGATCCATCATGAAAGATGATGTAATTCTCATTT 674
Qy 212 LeuTrpAsnGlyGluTrpLysArgTyrAsnGluAsnGlnTrpGluIleArgLysArgVal 231
Db 675 GCTGATGATACTGTGTATCGCAGAAATGGTCTATCCTCACAATAAATTAAGGAACCGAAT 734
Qy 232 TrpGluMetAlaAsnTyrValAsnMetLeuTyrLysLysLeuAsnThrHisValAlaLeu 251
Db 735 TGGGAATGTGCAATTTGTCAACATGATTTATATAAACCTTTAAACATCCATGTGACGTTG 794
Qy 252 ValGlyMetGluIleTrpThrTrpLysTrpLysIleLysIleThrProAsnAlaSerTrp 271
Db 795 GTTGGCATTGAATATGACACATGAAGATAAATAGAACATATATTCATATAGAAACT 854
Qy 272 ThrLeuGluAsnTrpSerLysTrpArgLysSerValLeuSerArgArgLysArgHisTrp 291
Db 855 ACCTTATTTGGGTTTTCATTTTGCAGAGAAAGATCCTTAAACACAGGAAGATTTTGAT 914
Qy 292 IleAlaGlnLeuIleThrAlaThrGluLeuAlaGlyThrThrValGlyLeuAlaTrpMet 311
Db 915 CATTGTTGATTTACTGAGGAACTGGCTCTACTCATCATGTGAGGAATTTCTTATCCA 974
Qy 312 SerThrMetTrpSerPro---TyrSerValGlyValValGlnTrpHisSerTrpAsnLeu 330
Db 975 GGGGGTATGCGCTGCTTATTTCCACCATGATCATTAAAG-----GATCTT 1022
Qy 331 LeuArg-----ValAlaGlyThrMetAlaHisGluMetGlyHisAsnTrpGly 346
Db 1023 TTACTCTGACACAAACATAATTTGAAACAGAAATGACATCAATCACTGGGCGCATACCTTGGG 1082
Qy 347 MetTrpHisTrpTrpTyrSerTrpLysTrpProSerThrIleTrpValMetTrpLysAla 366
Db 1083 ATGAGCATGACAGGTTCCCATGTCACCTGCTCTCAGAAAATTCGATGGACAGTGAT 1142
Qy 367 LeuSerTrpTyrIleProThrTrp---TrpSerSerTrpSerArgLeuSerTyrTrpLys 385
Db 1143 GGAAGC-----ATTCTGCACTGAATTCAGTAAATGCGCCAAACCAATACCAACAG 1196
Qy 386 TrpTrpGluTrpLysLeuSerAsnTrpLeuTrpAsnAlaProLeuProThrTrpIleIle 405
Db 1197 TACTTGAAGGATATTAAGCCACATGATGCTCAACATTCATTTCCCTTACAATTTTCAT 1256
Qy 406 SerThrProIleTrpGlyAsnGlnLeuValGluMetGlyGluTrpTrpTrpGlyThr 425
Db 1257 GATTTCGAATTTGTGGAACACAGAGTTGGATGAGGTGAGAGTGATGCTGCGGCCCT 1316
Qy 426 SerGluGluTrpThrAsnIleTrpTrpAlaLysThrTrpLysIleLysAlaThrTrp 445
Db 1317 GCTCAGAGTGTTACTTAATCTTGCTGTGATGCACACATGCTGACTGAGCCAGGATTT 1376
Qy 446 GlnTrpAlaLeuGlyGluTrpTrpGluLysTrpGlnTrpLysLysAlaGlyMetValTrp 465
Db 1377 ACTTGTGCAGAGAGAGAAATCTGTGAATCTTGTGAGATAAAGAGAGAGGCTCCATATGC 1436
Qy 466 ArgProAlaLysTrpGluTrpTrpLeuProGluMetTrpAsnGlyLysSerGlyAsnTrp 485
Db 1437 AGACCGCGGAAGATGAATGTGATTTCTTGAGATGTGCATGCGCCACTGCGCTGCTGT 1496

Qy 486 ProTrpTrpArgTrpGlnValAsnGlyTrpProTrpHisHisGlyLysGlyHisTrpLeu 505
Db 1497 CTTAAGGACCGATTCTGGGTCAATGGATTTCTTCCAGAACTCAGAAAGCTACTGTTTC 1556
Qy 506 MetGlyThrTrpProTrpLeuGlnGluGlnTrpThrGluLeuTrpGlyProGlyThrGlu 525
Db 1557 ATGGGAAATGTCCTCAACTCGTGAGGATCAGTCTCTGAACATATTTGATGATGAGGCAATA 1616
Qy 526 ValAlaTrpTrpLysSerTrpTyrAsnArgAsnGluGlyGlySerLysTyrGlyTyrTrpArg 545
Db 1617 GAGATCTCATGATATCTGCTACAGATGAATCAAAAAGGAAATAAATTTGGATCTGCATA 1676
Qy 546 ArgValTrpTrpThrLeuIleProTrpLysAlaAsnTrpThrMetTrpGlyLysLeuTrp 565
Db 1677 AACAGGAAAACAGATTCTTCCCTGTCGAGGAGAGATGTCAGATGTGGAAAGATCTAC 1736
Qy 566 TrpGlnGlyGly-----SerTrpAsnLeu----- 573
Db 1737 TGCATCTGAGGGGAGCTTCTCTCTCTGAGGAGAACAGACTTATCACCTTAAGGAT 1796
Qy 574 ProTrpLysGlyArgIleValThrTrpLeuThrTrpLysThrTrpTrpProGluTrpThr 593
Db 1797 CCCAGAGAAATGCTACTGTCT-----AAATGCAAAACATTTTTTTATACCATGAT 1847
Qy 594 SerGlnGluIleGlyMetValAlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpIle 613
Db 1848 TCTACAGACATGGCTGTGGCTGCGCTCAGAACAAATGTGGAGAGGAAATGATATGCAAC 1907
Qy 614 AsnAlaGluTrpValTrpIleGluLysAlaTyrLysSerThrAsnTrpSerSerLysTrp 633
Db 1908 AATGCTGAATCTTAAACATGGAAGAGTCTATATCTCAACCAATTGCCCTCTCAGTGC 1967
Qy 634 LysGlyHisAlaValTrpTrpHisGluLeuGlnTrpGlnTrpGluGluGlyTrpIlePro 653
Db 1968 AATGAAATCTGTAGTGGCCCGACTCCAGGCCACTGTGAGAGAGGAGCAGGCACCT 2027
Qy 654 ProTrpTrpTrpTrpSerSerValTrpHisTrpSerIleValValGlyValLeuTrp 673
Db 2028 GTAGCTGTGAAGAACCTTACATGTTTACCAGTATCACCATCTTGGTTGTGCTTGTCTC 2087
Qy 674 ProMetAlaValIleTrpValValAlaMetValIleArgHisGlnSerSerArgGlu 693
Db 2088 CTGTTTATTTGCGGTATCGGAGTCTTTATCTATTAGTTCGTACCGAAATGTATCAAG 2147
Qy 694 LysGlnLysLysTrpGlnArgProLeuSerThrThrGly----- 706
Db 2148 TTGAGCAAGTTTCAGAGCCCACTACAGAAACCTCGGAGTGGAGAACAAAGGATCTTT 2207
Qy 707 -----ThrArgPro-----HisLysGlnLys 713
Db 2208 GGTGATGAGCAGCAGAGATAAGGACTGAGCCATCTCGCCAGAAATTCATTTCCTAAATCAG 2267
Qy 714 ArgLysProGlnMetValLysAlaValGlnProGlnGluMetSerGlnMetLysProHis 733
Db 2268 AGAATCCAGAAATCTTGGAAAGCGCTG---CCCACTAGTTTTTCA-----AGTCCCAAC 2318
Qy 734 ValTyrTrpLeuProValGluGlyAsnGluProProAlaSerTrpHisLysTrpThrAsn 753
Db 2319 TACATCCACTG-----AAACCTGCAAGTAAGATTCAGAGGAAATC 2360
Qy 754 AlaLeuProProThrValTrpLysTrp 762
Db 2361 GCAGATCCCAATCAAAAGTGCCAAAGTGG 2387

Search completed: June 10, 2004, 03:04:54
Job time : 848 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 9, 2004, 21:07:21 ; Search time 139 Seconds

(without alignments)

3094.151 Million cell updates/sec

Title: US-10-726-148A-15

Perfect score: 4712

Sequence: 1 MLQGLLPVSLLSVAISAIAK.....PPTVKKNNPMSPKNSPKA 775

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10726148/rurat.09062004.161329.5304/app_query.fasta_1.967
-DB=Issued Patents NA -OPMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptn -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10726148 -CGN 1_1_69 @runat.09062004.161329.5304 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA.*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3128	66.4	2325	4	US-09-786-256C-16
2	3128	66.4	2325	4	US-09-786-256C-31
3	2132	45.2	1620	4	US-09-786-256C-3
4	2132	45.2	2560	4	US-09-786-256C-29
5	2098	44.5	1590	2	US-08-836-442-2
6	2098	44.5	2056	2	US-08-836-442-1
7	739	15.7	2335	4	US-09-026-001A-9
8	736	15.6	2359	4	US-09-026-001A-17
9	731.5	15.5	2050	4	US-09-026-001A-5
10	702	14.9	2209	1	US-08-514-014-1
11	702	14.9	2209	2	US-08-833-823-1
12	690	14.6	2280	3	US-08-813-150-1

13	590	14.6	2280	4	US-09-546-553-1	Sequence 1, Appli
14	677.5	14.4	2309	4	US-09-026-001A-13	Sequence 13, Appli
15	671.5	14.3	2297	4	US-09-026-001A-7	Sequence 7, Appli
16	662	14.0	2288	4	US-09-026-001A-11	Sequence 11, Appli
17	635	13.5	1851	4	US-09-608-790-2	Sequence 2, Appli
18	546.5	11.6	2830	2	US-09-010-928B-1	Sequence 1, Appli
19	528	11.2	2824	2	US-09-010-928B-3	Sequence 3, Appli
20	522	11.1	4847	4	US-10-164-595-57	Sequence 57, Appli
21	514	10.9	2824	4	US-07-757-022B-13	Sequence 13, Appli
22	514	10.9	3066	4	US-07-757-022B-83	Sequence 83, Appli
23	514	10.9	3117	4	US-07-757-022B-73	Sequence 73, Appli
24	514	10.9	3148	4	US-07-757-022B-57	Sequence 57, Appli
25	514	10.9	3420	4	US-07-757-022B-103	Sequence 103, App
26	514	10.9	3813	4	US-07-757-022B-43	Sequence 43, Appli
27	514	10.9	3936	4	US-07-757-022B-41	Sequence 41, Appli
28	514	10.9	3942	4	US-07-757-022B-141	Sequence 141, App
29	514	10.9	3945	4	US-07-757-022B-49	Sequence 49, Appli
30	514	10.9	3963	4	US-07-757-022B-45	Sequence 45, Appli
31	514	10.9	3963	4	US-07-757-022B-59	Sequence 59, Appli
32	514	10.9	4065	4	US-07-757-022B-47	Sequence 47, Appli
33	514	10.9	4086	4	US-07-757-022B-39	Sequence 39, Appli
34	514	10.9	4092	4	US-07-757-022B-51	Sequence 51, Appli
35	514	10.9	4215	4	US-07-757-022B-61	Sequence 61, Appli
36	514	10.9	5008	4	US-07-757-022B-1	Sequence 1, Appli
37	514	10.9	5041	4	US-09-023-655-981	Sequence 981, App
38	492.5	10.5	3431	4	US-09-632-098-1	Sequence 1, Appli
39	490.5	10.4	1392	3	US-09-411-329C-7	Sequence 7, Appli
40	490.5	10.4	1392	4	US-09-411-329C-5	Sequence 5, Appli
41	490.5	10.4	1392	4	US-09-466-276-5	Sequence 5, Appli
42	490.5	10.4	1392	4	US-09-846-729A-7	Sequence 7, Appli
43	490.5	10.4	1620	3	US-09-411-329C-12	Sequence 12, Appli
44	490.5	10.4	1620	3	US-09-411-329C-13	Sequence 13, Appli
45	490.5	10.4	1620	4	US-09-846-729A-12	Sequence 12, Appli

ALIGNMENTS

RESULT 1

US-09-786-256C-16
; Sequence 16, Application US/09786256C
; Patent No. 6680189
; GENERAL INFORMATION:
; APPLICANT: YOSHIMURA, Koji
; APPLICANT: HIKICHI, Yuichi
; APPLICANT: NISHIMURA, Atsushi
; TITLE OF INVENTION: No. 8680189el Protein and DNA Thereof
; FILE REFERENCE: 2544 USOP
; CURRENT APPLICATION NUMBER: US/09/786.256C
; CURRENT FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: PCT/JP99/04766
; PRIOR FILING DATE: 1999-09-02
; PRIOR APPLICATION NUMBER: JP 10-250115
; PRIOR FILING DATE: 1998-09-03
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 2325
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1) .. (2325)
; OTHER INFORMATION: Isolated nucleic acid encoding for a protein of SEQ ID NO. 15
US-09-786-256C-16

Alignment Scores:
Pred. No.: 3.22e-228 Length: 2325
Score: 3128.00 Matches: 654
Percent Similarity: 88.13% Conservative: 29
Best Local Similarity: 84.39% Mismatches: 92
Query Match: 66.38% Indels: 0
DB: 4 Gaps: 0

US-10-726-148A-15 (1-775) x US-09-786-256C-16 (1-2325)

QY 1 MetLeuGlnGlyLeuLeuProValSerLeuLeuLeuSerValAlaValSerAlaLeuLys 20
Db 1 ATGTTGCAAGGTCCTCGCAGTCAGTCTCCCTCTCTGTTGCGAGTAAGTCTATAAAA 60
QY 21 GluLeuProGlyValLysLysTyrGluValValTyrProIleArgLeuHisProLeuHis 40
Db 61 GAATCCCTCGGGGTGAAGAAGTATGAAGTGTTTATCTATATGAAGTCTCATCCATGCTAT 120
QY 41 LysArgGluAlaLysGluProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 60
Db 121 AAAAGAGAGGCCAAAGACGACAGACACAGAGAACCAATTGAACTCAATTAAGTATAAA 180
QY 61 MetThrIleAsnGlyLysIleAlaValLeuTyrLeuLysLysAsnLysAsnLeuLeuAla 80
Db 181 ATGACAAATTATGGAAAAATTGCAATGCTTTTATTTGMAAAAAAACAAGAACCTCTCTGCA 240
QY 81 ProGlyTyrThrGluThrTyrTyrAsnSerThrGlyLysGluIleThrThrSerProGln 100
Db 241 CCAGGCTACAGGAAACATATTATTAATTCACCTGGAAAGGAGATCACACAGGCCACAA 300
QY 101 IleMetTyrTyrTyrTyrTyrGlnGlyHisIleLeuAsnGluLysValSerThrAlaSer 120
Db 301 ATTATGGATGATGTATTATATCAAGGCATATTCTTAATGAAGAGGTTTCTGACGCTAGC 360
QY 121 IleSerThrTyrArgGlyLeuArgGlyTyrTyrSerGlnGlyTyrGluArgTyrTrpIle 140
Db 361 ATCAGCACATGTAGGGTCTAAGGGCTACTTCAGTCAGGGGGATCAAGATGACTTTATT 420
QY 141 GluProLeuSerProIleHisArgTyrpGlyGlnGlnHisAlaLeuTyrLysTyrAsnPro 160
Db 421 GAACCTTTAAGCCCATACATCGGATGAGCAGAGGATGACACTCTTCAAGTATATAACCT 480
QY 161 TrpGluLysAsnTyrTyrSerThrTyrpGlyMetTyrpGlyValLeuTyrAlaHisTyrLeu 180
Db 481 GATGAAGAAGATTATGACAGCACCTGTGGGATGGATGGTGTGTGTGGGGCCACCATTTG 540
QY 181 GlnGlnAsnIleAlaLeuProAlaThrLysLeuValLysLeuLysTyrArgLysValGln 200
Db 541 CAGCAGACATGTCCTACCTGCCACCAACTAGTAAATTTGAAGAAGCAGGAGGTTGAG 600
QY 201 GluHisGluLysTyrIleGluTyrTyrLeuValLeuTyrAsnGlyGluTyrLysArgTyr 220
Db 601 GAACATGAGAAATACATAGATAATTATTGTCCTGGATAATGTGTAGTTTAAAGGTAC 660
QY 221 AsnGluAsnGlnTyrpGluIleArgLysArgValTyrpGluMetAlaAsnTyrValAsnMet 240
Db 661 AATGAGAAATCAAGATGAGATCAGAAAGAGGGTATTTGAGATGGCTAATTAATGTCAACATG 720
QY 241 LeuTyrLysLysLeuAsnThrHisValAlaLeuValGlyMetGluIleTyrThrTyrLys 260
Db 721 CTTTATAAAGCTCAATACACTCACTGGCCTTAGTTGGTATGSAATCTGGACTGACAAAG 780
QY 261 TrpLysIleLysIleThrProAsnAlaSerThrThrLeuGluAsnTyrSerLysTyrArg 280
Db 781 GATAAGATAAAGATAAACCCTAAATGCAAGCTTCACCTTGGAGAAATTTCTTAATGGAGG 840
QY 281 GlySerValLeuSerArgArgLysArgHisTyrpIleAlaGlnLeuIleThrAlaThrGlu 300
Db 841 GGGAGTGTCTCTCAAGAAGAAAGCGTCATGATATGCTCAGTTAAATCACAGCAACAGAA 900
QY 301 LeuAlaGlyThrThrValGlyLeuAlaTyrMetSerThrMetTyrpSerProTyrSerVal 320
Db 901 CTGCTGGAAGACTGTGGGTCTTGCAATTATGATGTCACAAATGTTCTCTTATCTGTT 960
QY 321 GlyValValGlnTyrHisSerThrAsnLeuLeuArgValAlaGlyThrMetAlaHisGlu 340
Db 961 GGGCTGTCTCAGGACCAACAGCGATAATCTCTTAGAGTTGACAGGACAAATGGCACATGAA 1020
QY 341 MetGlyHisAsnTyrpGlyMetThrHisTyrTyrTyrSerTyrpLysTyrpProSerThrIle 360
Db 1021 ATGGGCCAACAACTTTGGAATGTTTCAATGACGACTATTCTTGCAGAGTGTCTTCTTACAATA 1080

QY 361 TrpValMetTyrpLysAlaLeuSerTyrTyrIleProThrTyrTyrpSerTyrpSerArg 380
Db 1081 TGTGTATGAGCAAAAGCACTGAGCTTCTATATATACCCACAGACTCATGTTCTCTCAGCCGT 1140
QY 381 LeuSerTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyrTyr 400
Db 1141 CTCAGCTATGACAGATTTTTCAGATAAATTAATTAATTAATTAATTAATTAATTAATTAAT 1200
QY 401 ProThrTyrpIleIleSerThrProIleTyrpGlyAsnGlnLeuValGluMetGlyGluTyr 420
Db 1201 CCTACAGATATCATATCCACTCCAATTTGTGGGAACCACTGCTGGAAATGGAGAGGAC 1260
QY 421 TrpTyrTyrGlyThrSerGluGluTyrThrAsnIleTyrTyrTyrpAlaLysThrTyrLys 440
Db 1261 TGTGATTTGGGACATCTGAGGAATGTATCCAAATTTTGTCTGTGATGCTTAAGCATGTAAA 1320
QY 441 IleLysAlaThrTyrpGlnTyrpAlaLeuGlyGluTyrTyrpGluLysTyrpGlnTyrLys 460
Db 1321 ATCAAGCAACTTTTCAATGTGATAGGAGATGTTGTGAAAAATGCCAATTTAAAGAG 1380
QY 461 AlaGlyMetValTyrpArgProAlaLysTyrpGluTyrTyrpLeuProGluMetTyrpAsnGly 480
Db 1381 GCTGGATGGTGTGCAGACCAAGAAAGATGAGTGCACCTGCCCTGAATGTGTAAATGTT 1440
QY 481 LysSerGlyAsnTyrpProTyrTyrpArgTyrpGlnValAsnGlyTyrpProTyrpHisGly 500
Db 1441 AAATCTGTAATTTGCTGATGATAGATTCCAAGTCAATGGCTTCCCTGCCATCACGGG 1500
QY 501 LysGlyHisTyrpLeuMetGlyThrTyrpProThrLeuGlnGlnTyrpThrGluLeuTyr 520
Db 1501 AAGGCCCACTGCTTGTATGGGACATGCCACACATGCGAGGACAGTGCACAGCTGTGG 1560
QY 521 GlyProGlyThrGluValAlaTyrpLysSerTyrpAsnArgAsnGluGlyGlySerLys 540
Db 1561 GGACAGGAACCTGAGGTTGCAGATAAGTCATGTTACAAACAGGAATGAAGTGGTGGTCAAAG 1620
QY 541 TyrGlyTyrTyrpArgArgValTyrpThrLeuIleProTyrpLysAlaAsnTyrpMet 560
Db 1621 TACGGTACTGTGCGAGAGTGGATGACACACTCACTTCCCTGCAAGCAAAATGATACCATG 1680
QY 561 TrpGlyLysLeuTyrpTyrpGlnGlyLysSerTyrpAsnLeuProTyrpLysGlyArgIleVal 580
Db 1681 TGTGGAGTTGTTCTGTCAAGTGGGTCGGTAATTTGCCCTGCAAGGACGGATAGT 1740
QY 581 ThrTyrpLeuThrTyrpLysThrTyrpProGluTyrpThrSerGlnGluIleGlyMetVal 600
Db 1741 ACTTTCCTGACATGTAAACATTTGATCCTGAAAGACACAAAGTCAAGAAATAGGCATGGT 1800
QY 601 AlaAsnGlyThrLysTyrpGlyTyrpAsnLysValTyrpIleAsnAlaGluTyrpValTyrp 620
Db 1801 GCCAATGCACTAGTGTGGCGATTAACAAAGTTTGCATTAATGCAAGATGTGTGATATT 1860
QY 621 GluLysAlaTyrpLysSerThrAsnTyrpSerSerLysTyrpLysGlyHisAlaValTyrp 640
Db 1861 GAGAAAGCCTACAATCAACCAATTTGCTCATCTAAGTCAAAAGGACATGCTGTGTGTAC 1920
QY 641 HisGluLeuGlnTyrpGlnTyrpGluGluGlyTyrpIleProProTyrpTyrpTyrpSerSer 660
Db 1921 CATGAGTCCAGTGTCAATGTGAGGAAGGATGGATCCCTCCCGACTGCGATGACTCTCA 1980
QY 661 ValValTyrpHisTyrpSerIleValValGlyValLeuTyrpProMetAlaValIleTyrpVal 680
Db 1981 GTGCTCTCCACTCTCCATTTGTGTTGGGTTGTGTTGCCAATGGCGGTCTATTTTGTG 2040
QY 681 ValValAlaMetValIleArgHisGlnSerSerArgGluLysGlnLysLysTyrpGlnArg 700
Db 2041 GTGGTTCTATGTTAAATCCGGCACACAGAGTCCAGAGAAAGCAGAGAAAGATCAGAGG 2100
QY 701 ProLeuSerThrThrGlyThrArgProHisLysGlnLysArgLysProGlnMetValLys 720
Db 2101 CCACATCTACCACTGGCCAGGCCACACAAAACAGAAAGGAAACCCCGATGTTAAAG 2160

Db 61 GAACCTCCCTGGGGTGAAGAAGTATGAGTGGCTTTATCTCTATAAGACTTCATCCACTGCAT 120
Qy 41 LysArgGluAlaLysGluProGluGlnGlnGlnTrpGluThrGluLeuLysTyrLys 60
Db 121 AAAAGAGAGGCCAAAGAGCCAGACCAACAGAACAAATTTGAACCTGGAATTAAAGTATATAA 180
Qy 61 MetThrIleAsnGlyLysIleAlaValLeuTyrLysLysAsnLysAsnLeuLeuAla 80
Db 181 ATGACAATTATGAAAATTCAGTGCTTTATTGAAAATAAACAAGAACTCTTGCA 240
Qy 81 ProGlyTyrThrGluThrTyrAsnSerThrGlyLysGluIleThrThrSerProGln 100
Db 241 CCAGCTACACAGAAACATATTATAATCCACTGGAAGAGAGATCACCACAGGCCACAA 300
Qy 101 IleMetTrpTrpTrpTyrGlyGlyHisIleLeuAsnGluLysValSerTrpAlaSer 120
Db 301 ATTATGGATGATGCTTATTATCAGGACATATCTTAATGAAGAGTTTCTGAGCTAGC 360
Qy 121 TieserThrTrpArgGlyLeuArgGlyTyrTrpSerGlnGlyTyrGlnArgTyrTrpIle 140
Db 361 ATCAGCACATCTAGGGTCTAAGGGCTACTTCAGTCAGGGGATCAAGATACCTTATT 420
Qy 141 GlnProLeuSerProIleHisArgTrpGlyGlnGluHisAlaLeuTrpLysTyrAsnPro 160
Db 421 GAACTTTTAAGCCCATACATCGGGATGGACAGGAGCATGCACCTCTCAAGTATAACCT 480
Qy 161 TrpGluLysAsnTrpTrpSerThrTrpGlyMetTrpGlyValLeuTrpAlaHisTrpLeu 180
Db 481 GATGAAGAATATGACAGCACCTGTGGATGATGATGGTGTGTGGGCCACGATTG 540
Qy 181 GlnGlnAsnIleAlaLeuProAlaThrLysLeuValLysLeuLysTrpArgLysValGln 200
Db 541 CAGCAGAACATTCCTACCTGCCACCAACCTAGTAAATTTGAAGACAGGAGGTTCCAG 600
Qy 201 GluHisGluLysTyrIleGluTyrTrpLeuValLeuTrpAsnGlyGluTrpLysArgTyr 220
Db 601 GAACATGAGAAATACATAGATATTATTGTGCTCGGATAATGTTGAGCTTTAAAGGTAC 660
Qy 221 AsnGluAsnGlnTrpGluIleArgLysArgValTrpGluMetAlaAsnTyrValAsnMet 240
Db 661 AATCAGAAATCAGATGAGATCAGAAAGAGGATTTGAGATGGCTGAATATGTCACATG 720
Qy 241 LeuTyrLysLysLeuAsnThrHisValAlaLeuValGlyMetGluIleTrpThrTrpLys 260
Db 721 CTTTATAAAGCTCACTACTCATGTCGCTTAGTTGGTATGTAATCTGGACTGACAG 780
Qy 261 TrpLysIleLysIleThrProAsnAlaSerTrpThrLeuGluAsnTrpSerLysTrpArg 280
Db 781 GATAAGATAAAGATAACCCCAATCAAGCTTCACCTTGGAGAAATTTCTAAATGGAGG 840
Qy 281 GlySerValLeuSerArgArgLysArgHisTrpIleAlaGlnLeuIleThrAlaThrGlu 300
Db 841 GGGAGTGTCTCTCAAGAAGAAAGCGTATGATATGCTCAGTTAATCCACAAACAGAA 900
Qy 301 LeuAlaGlyThrThrValGlyLeuAlaTrpMetSerThrMetTrpSerProTyrSerVal 320
Db 901 CTGTCTGGAACACTGTGGTCTCTCAATTATGTCTACAAATGTTCTCCCTTATCTGTT 960
Qy 321 GlyValValGlnTrpHisSerTrpAsnLeuLeuArgValAlaGlyThrMetAlaHisGlu 340
Db 961 GCGGTGTTCAGGACACAGCGATATCTCTTAGAGTTGCGAGGACAAATGCGACATGAA 1020
Qy 341 MetGlyHisAsnTrpGlyMetTrpHisTrpTrpTyrSerTrpLysTrpProSerThrIle 360
Db 1021 ATGGGCCCAACATTTGGAAATGTTTCATGACGACTATCTGCAAGTGTCCTTCTACAATA 1080
Qy 361 TrpValMetTrpLysAlaLeuSerTrpTyrIleProThrTrpTrpSerSerTrpSerArg 380
Db 1081 TGTGTGATGGAACAAAGCACTGAGCTCTATATACCCACAGACTTCAGTTCTCTGAGCGCT 1140
Qy 381 LeuSerTrpTrpLysTrpGluTrpLysLeuSerAsnTrpLeuTrpAsnAlaProLeu 400
Db 1141 CTCAGCTATGACAGCTTTTGAAGATAAAATATCAAAATTCCTTCTCTTAAATCTCCATTG 1200

Qy 401 ProThrTrpIleIleSerThrProIleTrpGlyAsnGlnLeuValGluMetGlyGluTrp 420
Db 1201 CCTACAGATATCATATCCACTCCAAATTTGTGGAAACCAAGTTGTTGAAATGGAGAGGAC 1260
Qy 421 TrpTrpTrpGlyThrSerGluGluTrpThrAsnIleTrpTrpTrpAlaLysThrTrpLys 440
Db 1261 TGTGATTGGGACATCTGAG-----ACATGTAAA 1290
Qy 441 IleLysAlaThrTrpGlnTrpAlaLeuGlyGluTrpTrpGluLysTrpGlnTrpLysLys 460
Db 1291 ATCAAAGCACTTTTCAATGTGCATTAGAGAAATGTTGTGAAAAATGCCAATTTAAAAAG 1350
Qy 461 AlaGlyMetValTrpArgProAlaLysTrpGluTrpTrpLeuProGluMetTrpAsnGly 480
Db 1351 GCTGGGATGGTGTGCAGACCCAGCAAGATGATGATGCGACCTGCTGAAATGTGTAATGT 1410
Qy 481 LysSerGlyAsnTrpProTrpTrpArgTrpGlnValAsnGlyTrpProTrpHisGly 500
Db 1411 AAATCTGTAATGTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1470
Qy 501 LysGlyHisTrpLeuMetGlyThrTrpProThrLeuGlnGluGlnTrpThrGluLeuTrp 520
Db 1471 AAGGCCCACTGCTGATGCGGACATGCCCCACATGCGAGGACGTCGACAGCTGTGG 1530
Qy 521 GlyProGly 523
Db 1531 GGACCAAGT 1539

RESULT 6
US-08-836-442-1
; Sequence 1, Application US/08836442
; Patent No. 5990293
; GENERAL INFORMATION:
; APPLICANT: DOCHERTY, Andrew, J.P.
; APPLICANT: SLOCUMBE, Patrick, M.
; TITLE OF INVENTION: A HUMAN METALLOPROTEINASE
; TITLE OF INVENTION: VARIANTS THEREOF AND DNA SEQUENCES CODING THEREFOR
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Fast-SEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,442
; FILING DATE: 01-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB96/02181
; FILING DATE: 13-MAR-1997
; APPLICATION NUMBER: GB 9612150.4
; FILING DATE: 11-JUN-1996
; APPLICATION NUMBER: GB 9526229.1
; FILING DATE: 21-DEC-1995
; APPLICATION NUMBER: GB 9521498.7
; FILING DATE: 20-OCT-1995
; APPLICATION NUMBER: GB 95521495.3
; FILING DATE: 20-OCT-1995
; APPLICATION NUMBER: GB 9518023.8
; FILING DATE: 05-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Resnick, David S
; REGISTRATION NUMBER: 34,235
; REFERENCE/DOCKET NUMBER: 47425
; TELECOMMUNICATION INFORMATION:

COUNTRY: USA
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/026,001A
 FILING DATE: 18-FEB-1998
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Brown, Scott A.
 REGISTRATION NUMBER: 32,724
 REFERENCE/DOCKET NUMBER: G15293B
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 498-8224
 TELEFAX: (617) 876-5851
 INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2335 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-09-026-001A-9

Alignment Scores:
 Pred. No.: 5,37e-47 Length: 2335
 Score: 739.00 Matches: 217
 Percent Similarity: 51.01% Conservative: 110
 Best Local Similarity: 33.85% Mismatches: 282
 Query Match: 15.68% Indels: 32
 DB: 4 Gaps: 11

US-10-726-148A-15 (1-775) x US-09-026-001A-9 (1-2335)

QY	177	AlaHisTrpLeuGlnAsnIleAlaLeuProAlaThrLysLeuValLysLeu---	lys	195
Db	595	ACTACTTTGGGAGTCAGATGAGTCCATTGAAAGACCTCTCAGTTAACTAACACCTCTCTGAA	---	654
QY	196	TrpArgLysValGlnGluHisGluLysTyrIleGluTyrTyrLeuValTrpAsnGly	215	
Db	655	CAAGACAGGTACTTGCAGGCCAAAATACTCGAGTTTACGTTTGTGGACAACTA	714	
QY	216	GluTrpLysArgTyrAsnGluAsnGlnTrpGluLeuArgLysArgValTrpGluMetAla	235	
Db	715	ATGTACAGGCATTACAAACCGGATAAACCTGTTATAAAAGACAGTATATGAAATGATC	774	
QY	236	AsnTyrValAsnMetLeuTyrLysLysLeuAsnThrHisValAlaLeuValGlyMetGlu	255	
Db	775	AACACTATGAATATGGTGTACATCGTTTGAATTTTACATAGCACTGATTGGGCTAGAA	834	
QY	256	IleTrpThrTrpLysTrpLysIleLysIleThrProAsnAlaSerTrpThrLeuGluAsn	275	
Db	835	ATTGTGTCACACAGAAATGAGATTAAATGTCATCAGACGTCAGGCCACTTTGGACTTA	894	
QY	276	TrpSerLysTrpArgGlySerValLeuSerArgLysArgHisTrpIleAlaGlnLeu	295	
Db	895	TTTGGAGATCGAGAGAAAAAATTCCTGCCACGCAAAAGGAATGATATGCTCAGTTA	954	
QY	296	IleThrAlaThrGluLeuAlaGlyThrValGlyLeuAlaTrpMetSerThrMetTrp	315	
Db	955	CTCAGGGTATTGACTTCAAGGAACCTCTGTAGGACTTGCCTTACATAGGTTCCATCTGC	1014	
QY	316	SerPro---TyrSerValGlyValValGlnTrpHisSerTrpAsnLeuLeuArgValAla	334	
Db	1015	AATCGAAGAGTCTCTGTAGCAGTTGTTTCAGATTATAGCAGTAGAACACAGCATCGTGGCA	1074	
QY	335	GlyThrMetAlaHisGluMetGlyHisAsnTrpGlyMetTrpHisTrpTrpLysSerTrp	354	
Db	1075	ATTACAATGGCCCATGAGATGGGTGCTATATATGGGCACTCATATGACGACGACCTTCCTGT	1134	
QY	355	LysTrpProSerThrIleTrpValMetTrpLysAlaLeuSerTrpTyrIleProThrTrp	374	
Db	1135	ACTTGTGGTTCACAAATGCGTTATG-----TCTACAAGACGTACTGAACCTGCGCTAT	1198	
QY	375	---TrpSerSerTrpSerArgLeuSerTyrTrpLysTrpTrpGluTrpLysLeuSerAsn	393	
Db	1189	CAGTTCAGCTCTGTAGTTCGCGGAAACATCAGAGATATCTTCTTAGACAGACACCACAA	1248	
QY	394	TrpLeuTrpAsnAlaProLeuProThrTrpIleIleSerThrProIleTrpGlyAsnGln	413	
Db	1249	TGCATTCTCAACAAACCTTGACACAGATATTTGTTTTCACCTCCAAATTTGTGGAATAAC	1308	
QY	414	LeuValGluMetGlyGluTrpTrpTrpGlyThrSerGluGluTrpThrAsnIleTrp	433	
Db	1309	TTTGTGGAGTGGAGAGAAATGTGACTGTGGCTCTCTCTGCGGATTTGTCAAAGTGGCTGC	1368	
QY	434	TrpTrpAlaLysThrTrpLysIleLysAlaThrTrpGlnTrpAlaLeuGlyGluTrpTrp	453	
Db	1369	TGCGACGCTACAACTGTAACTACAACTCATGACAGCTGACTCCGGAAGGTTGTGT	1428	
QY	454	GluLysTrpGlnTrpLysLysAlaGlyMetValTrpArgProAlaLysTrpGluTrpTrp	473	
Db	1429	GAGAAATGCAATTTAAGGGACGAGCAGAAATGCGGGGCGAGCAAGATGACTGTGAC	1488	
QY	474	LeuProGluMetTrpAsnGlyLysSerGlyAsnTrpProTrpTrpArgTrpGlnValAsn	493	
Db	1489	TTGCTTGAACCTGCACTGGCCATCTGCTGAGTGTCTCCACAGACATCTTCAGAGGAAT	1548	
QY	494	GlyTrpProTrpHisIleGlyLysGlyHisTrpLeuMetGlyThrTrpProThrLeuGln	513	
Db	1549	GGACTTCCATGCCAAAACAAACAGAGTTACTCTCAATGGGAAATGCCCATCATGACA	1608	
QY	514	GluGlnTrpThrGluLeuTrpGlyProGlyThrGluValAlaTrpLysSerTrpTrpAsn	533	
Db	1609	AACCAATGATTGCTCTCCGGGACGAGGTGTAAAGATGATCTCGAGATAGTGTGTTTACA	1668	
QY	534	ArgAsnGluGlyGlySerLysTyrGlyTyrTrpArgArgValTrpTrpThrLeuIlePro	553	

Db 1669 TTGAACACAGAGAACCAAGTGGTGTGGCTGTGGAGATGGAATGCTAGAAAGATTCCA 1728
QY 554 TrpLysAlaAsnTrpThrMetTrpGlyLysLeuTrpTrpGlnGlyGlySer 570
Db 1729 TGTGACAGCAAGAGATGTAAGTGTGGCAGTTATTTGCAAAAGGGAACTCGATGATA 1788
QY 571 TrpAsnLeuProTrpLysGlyArgIleValThrTrpLeuThrTrpLysThrTrpPro 590
Db 1789 TGCACCTGCTCAGTTCACACAGT 1812
QY 591 GluTrpThrSerGlnGluIleGlyMetValAlaAsnGlyThrLysTrpGlyTrpAsnLys 610
Db 1813 -----GACCCAGTTATGAATGGTGTGAACCTGGACAAATGTGGAGATGAAATG 1863
QY 611 ValTrpIleAsnAlaGluTrpValTrpIleGluLysAlaTyr---LysSerThrAsnTrp 629
Db 1864 GTGTGCAGCAACAGGCGAGTGTGTGATGTGAAGACAGCCTACTGATCAAGCACTGGCTTC 1923
QY 630 Ser 630
Db 1924 TCT 1926

RESULT 8

US-09-026-001A-17
; Sequence 17, Application US/09026001A
; Patent No. 6413760
; GENERAL INFORMATION:
; APPLICANT: Boodhoo, Amechand
; APPLICANT: Seehra, Jasbir
; APPLICANT: Shaw, Gray
; APPLICANT: Sako, Dianne
; TITLE OF INVENTION: HIGHLY PURIFIED MOCARHAGIN, A COBRA VENOM
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/09/026,001A
; APPLICATION NUMBER: US/09/026,001A
; FILING DATE: 18-FEB-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15293B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2359 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-026-001A-17
Alignment Scores:
Pred. No.: 1.59e-46 Length: 2359
Score: 736.00 Matches: 219
Percent Similarity: 50.38% Conservative: 109
Best Local Similarity: 33.64% Mismatches: 279
Query Match: 15.62% Indels: 44

DB: 4 Gaps: 12
US-10-726-148A-15 (1-775) x US-09-026-001A-17 (1-2359)
QY 1 MetLeuGlnGlyLeuLeuProValSerLeuLeuLeuSerVal-----AlaVal 16
Db 67 ATGATCCAAGCTCTCTTG---GTAGCTATATGCTTAGCGGTTTTTCATCATCAAGGGAGC 123
QY 17 SerAlaIleLysGluLeuProGlyValLysLysLysLysLysLysLysLysLysLysLys 36
Db 124 TCTATAATCCTGGAATCCGGGAATGTTAATGATTATGAAGTAGTGTATCCAAAAAGTG 183
QY 37 HisProLeuHisLysArgGluAlaLysGluProGlnGlnGlnGlnGlnGlnGlnGlnGln 56
Db 184 CTGTCATTTCCAAAGAGGAGGTTCAAGATCCTCAGCCAGACAGCAAGATGATGAATACA 243
QY 57 LeuLysTyrLysMetThrIleAsnGlyLysLysLysLysLysLysLysLysLysLysLys 76
Db 244 ATGCATATGAATTTCAAGTGAATGAGAGCCAGTAGTCTCTCACCAGAAAGAAATAAA 323
QY 77 AsnLeuLeuAlaProGlyTyrThrGluThrTyrTyrAsnSerThrGlyLysGluIleThr 96
Db 304 GGACTTTTTTCAGAAAGATTACACTGAACCTCATTATGCTCTGATGGCAGAGAAATTACA 363
QY 97 ThrSerProGlnIleMetTrpTrpTrpTyrTyrGlnGlyHisIleLeuAsnGluLysVal 116
Db 364 ACAAGCCCACTCGTTCAGGATCAGTCTATTATCATGTTTACATTGAGATGAAGCTGAC 423
QY 117 SerTrpAlaSerIleSerThrTrpArgGlyLeuArgGlyTyrTrpSerGlnGlyTrpGln 136
Db 424 TCAAGTCAGTCATCAGTCGATCGATCGCTTGAAGAGGACATTCGAGCTCAAGGGGAG 483
QY 137 ArgTyrTrpIleGluProLeuSerProIleHisArgTrpGlyGlnGluHisAlaLeuTrp 156
Db 484 ACATACITTTATTGAACCTTGAAGATTCCGAC-----AGTGAAGCCCATGCAATCTAC 537
QY 157 LysTyrAsnProTrpGluLysAsnTyrTrpSerThrTrpGlyMetTrpGlyValLeuTrp 176
Db 538 AAAGATGAAATGTAGAA---AACGAGGATGAGACCCCGGAAACCTGTGGGTAAACCGAG 594
QY 177 AlaHisTrp-----LeuGlnGlnAsnIleAla 185
Db 595 ACTACTTGGGAGTCAGATGAGTCCATTGAAAGACCTCTCAGTTAGACAGCAGCAGCAAG 654
QY 186 LeuProAlaThrLysLeuValLysLeuLysTrpArgLysValGlnGluHisLysLysTyr 205
Db 655 CGGCGCGCAACTAACACTCCTGAAACAGAC-----AGGTACTTCAGGCGCAAAAAATAC 708
QY 206 IleGluTyrTyrLeuValLeuTrpAsnGlyGluTrpLysArgTyrAsnGluAsnGlnTrp 225
Db 709 CTCGAGTTTACGTGGTGTGGACCAACATAATGACAGGCATTACAAACGCGATAAACCT 768
QY 226 GluIleArgLysArgValTrpGluMetAlaAsnTyrValAsnMetLeuTyrLysLysLeu 245
Db 769 GTTATAAAAGAGAGATATGAAATGATCACTACATGATGATATGTTGCTACATCGTTTG 828
QY 246 AsnThrHisValAlaLeuValGlyMetGluIleTrpThrTrpLysTrpLysLysLysLys 265
Db 829 AATTTTCATAGTACTGATTTGGCTAGAAATTTGGTCCACAGAAATGAGATTAAATGTG 898
QY 266 ThrProAsnAlaSerTrpThrLeuGluAsnTrpSerLysTrpArgGlySerValLeuSer 285
Db 883 CAATCAGCGTGGCGGCACTTTGGACTTATTTGGAGATGGAGAGAGAAAAAATITCTG 948
QY 286 ArgArgLysArgHisTrpIleAlaGlnLeuIleThrAlaThrGluLeuAlaGlyThrThr 305
Db 949 CCACGAAAAGGAATGATAATGCTCAGTTACTCAGCGGTATTGATCTCAAGGAGACCTCT 1008
QY 306 ValGlyLeuAlaTrpMetSerThrMetTrpSerPro---TyrSerValGlyValValGln 324
Db 1009 GTAGGACTTGCTTACATAGTCTCCATCTGCAATCCGAAGAGATTCTGTAGCAGTTGTT 1068
QY 325 TrpHisSerTrpAsnLeuLeuArgValAlaGlyThrMetAlaHisGluMetGlyHisAsn 344


```
Db 1069 GATTATAGCAGTAGAACAAGCAGTGGTGGCAATTAACAATGCCCATGAGATGGGTCAATAAT 1128
Qy 345 TrpGlyMetTrpHisTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrp 364
Db 1129 ATGGGCAATCATCATGAGGACCTCTCTGTACTTGTGGTCTTAACAAATGCGTTATG--- 1185
Qy 365 LysAlaLeuSerTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrp 383
Db 1186 ---TCATACAGAGAGTACTGAACCTGCTATCACTTGTAGTGTCCGGACAT 1242
Qy 384 TrpLysTrpTrpGluTrpLysLeuSerAsnTrpLeuTrpAsnAlaProLeuProTrpTrp 403
Db 1243 CAGGAGTATCTTCTAGACAGACACCAATGCAATCTCAACAAACCCCTTGACACAGAT 1302
Qy 404 IleLeuSerThrProLeuTrpGlyAsnGlnLeuValGluMetGlyGluTrpTrpTrpTrp 423
Db 1303 ATGTTTACCTCCCAATTTGGGAATAACTTTGGAGGTGGGAGGAAGATGTACTGT 1362
Qy 424 GlyThrSerGluGluTrpThrAsnIleTrpTrpTrpAlaLysThrTrpLysIleLysAla 443
Db 1363 GGCTCTCCGCGGATGTCAAGTGGCTGCTGGAGCCTACACCTGTAACTACAACT 1422
Qy 444 ThrTrpGlnTrpAlaLeuGlyGluTrpTrpGluLysTrpGlnTrpLysLysAlaGlyMet 463
Db 1423 CATGCACAGTGTGACTCCGAGGGTGTCTGAGAAATGCAAACTTAAGGAGCAGGAGCA 1482
Qy 464 ValTrpArgProAlaLysTrpGluTrpTrpLeuProGluMetTrpAsnGlyLysSerGly 483
Db 1483 GAATCCGGGAGCAAGAGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1542
Qy 484 AsnTrpProTrpTrpArgTrpGlnValAsnGlyTrpProTrpHisGlyLysGlyHis 503
Db 1543 GAGTGTCACACAGACATCTTCAGAGGATGCTTCATCCATGCGCAACCAACAGAGTTAC 1602
Qy 504 TrpLeuMetGlyTrpTrpProTrpLeuGlnGluTrpTrpGluLeuTrpGlyProGly 523
Db 1603 TGCTCAATGGGAAATGGCCCATCATGACAAACCAATGATTTGCTCTCCGGGACCAGGT 1662
Qy 524 ThrGluValAlaTrpLysSerTrpTrpAsnArgAsnGluGlyGlySerLysTrpGlyTrp 543
Db 1663 GTAAAGTATCTCGAGATAGCTGTGTATCATTTGAACCCAGAGAACCAAGTGTGGCTTG 1722
Qy 544 TrpArgArgValTrpTrpTrpLeuLeuProTrpLysAlaAsnTrpTrpMetTrpGlyLys 563
Db 1723 TGCAGATGGAATGGTAGAAGATTCATGTCAGCAAGAGATGTAAGTGTGGCAGG 1782
Qy 564 LeuTrpTrpGlnGlySer-----TrpAsnLeuProTrpLysGlyArgIleVal 580
Db 1783 TTATTTTCAAAAAGGGAACCTCGATGATATGCAACTGCTCAGTTTCCACCACT 1836
Qy 581 ThrTrpLeuTrpTrpLysTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrp 600
Db 1837 -----GACCCCAAGTTATGGAATGGTT 1857
Qy 601 AlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpIleAsnAlaGluTrpValTrpIle 620
Db 1858 GAACCTGGAAACAAATGCGAGATGGAATGGTGTGTCACCAACAGGACACTGTGTGATCTG 1917
Qy 621 GluLysAlaTrp---LysSerThrAsnTrpSer 630
Db 1918 AAGACAGCTTACTGATCAAGCACTGGCTTCTCT 1950
```

RESULT 9

```
US-09-026-001A-5
; Sequence 5, Application US/09026001A
; Patent No. 6413760
; GENERAL INFORMATION:
; APPLICANT: Boodhoo, Anechand
; APPLICANT: Seehra, Jasbir
; APPLICANT: Shaw, Gray
; APPLICANT: Sako, Dianne
; TITLE OF INVENTION: HIGHLY PURIFIED MOCARHAGIN, A COBRA VENOM
```

```
; TITLE OF INVENTION: THERAPEUTIC USES THEREOF
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/026,001A
; FILING DATE: 18-FEB-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15293B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2050 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 78..1940
; US-09-026-001A-5
Alignment Scores:
Pred. No.: 1,99e-46 Length: 2050
Score: 731.50 Matches: 216
Percent Similarity: 50.00% Conservative: 106
Best Local Similarity: 33.54% Mismatches: 259
Query Match: 15.52% Indels: 64
DB: 4 Gaps: 15
US-10-726-148A-15 (1-775) x US-09-026-001A-5 (1-2050)
Qy 1 MetLeuGlnGlyLeuLeuProValSerLeuLeuSerVal-----AlaVal 16
Db 78 ATGATCCAGGCTCTCTG---GTAGCTATATCTTAGCGGTTTTTCCATATCAAGGGAGC 134
Qy 17 SerAlaIleLysGluLeuProGlyValLysLysTrpGluValValTrpProIleArgLeu 36
Db 135 TCTATAATCCTGGAATCCGGAATCTTAATGATTATGAAGTAGTGATATCCACAAAAGTG 194
Qy 37 HisProIleHisLysArgGluAlaLysGluProGluGlnGlnGluTrpGluTrpGlu 56
Db 195 CTGCAITTTCCAAAGGAGGAGTTCAGAACTCCTCCAGCAGACCAAGTATGAAGATACA 254
Qy 57 LeuLysTrpLysMetThrIleAsnGlyLysIleAlaValLeuTrpLeuLysLysAsnLys 76
Db 255 ATGCAATATGATTTTCAGTGAACGAGAGCGAGTGTCTTCTTACCTTGAAGAAGATAAA 314
Qy 77 AsnLeuLeuAlaProGlyTrpThrGluTrpTrpTrpTrpTrpTrpTrpTrpTrpTrp 96
Db 315 GGACTTTTTTCAGAGAGATTACACTGAAACTCATTTATGCCCCCTGATGGCAGAGAAATTACA 374
Qy 97 ThrSerProGlnIleMetTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrpTrp 116
Db 375 ACAAGCTCTCCAGTTCCAGGATCATCTGCTATTATCATGCTTACATTACAGAACTGAC 434
Qy 117 SerTrpAlaSerIleSerThrTrpArgGlyLeuArgGlyTrpTrpSerGlnGlyTrpGln 136
Db 117 SerTrpAlaSerIleSerThrTrpArgGlyLeuArgGlyTrpTrpSerGlnGlyTrpGln 136
```

Db	435	TC	AGTGCAGTCATCAGTGCATGCTGATGGCTTGAAGACGACATTTCAAGCATCAAGCGGAG	494
Qy	137	Arg	TyTrrTpIleGluProLeuSerProIleHisArgTrpGlyGlnGluHisAlaLeuTrp	156
Db	495	AC	ATACTTTATTGAGCCCTTGGAGCTTCTGCAC-----AGTGAAGCCCATGCAATATAC	548
Qy	157	Lys	TyAsnProTrpGluLysAsnTyTrrSerThrTrpGlyMetTrpGlyLeuValLeuTrp	176
Db	549	AA	GATGAAATGTAGAGAGAGAGAA---GAGATCCCAAAATCTGTGGGTATACCAG	605
Qy	177	Ala	HisTrpLeuGlnGlnAsnIleAlaLeuProAlaThrLysLeuValLysLeuValTrp	196
Db	606	AC	TACTTGGGAATCAGATGAG-----CCGATTGAAAGTCTCTCAGTTAACTTAACT	656
Qy	197	Arg	LysValGlnGlnHis-----GluLysTyrrIleGluTyTrrLeuValLeu	212
Db	657	AC	TCTGAACAAGACAGGTACTTGCAGCCAAAAAATACATCGAGTTTTTACGTGGTTGTG	716
Qy	213	Trp	AsnGlyGluTrpLysArgTyAsnGluAsnGlnTrpGluIleArgLysArgValTrp	232
Db	717	GAC	ATGTAATGTACMGRAATACACCGGCAAGTTACATGTTATACAGAGAGATATAT	776
Qy	233	Glu	MetAlaAsnTyrrValAsnMetLeuTyrrLysLysLeuAsnThrHisValAlaLeuVal	252
Db	777	CAA	ATGGTCAACGGCTTTAAATACCATGTACAGAGCTTTGAATTTTCATAGCACATGATT	836
Qy	253	Gly	MetGluIleTrpThrTrpLysTrpLysIleLysIleThrProAsnAlaSerTrpThr	272
Db	837	GG	CTAGAAATTTGGTCAACGGAAATAGAGATTAATGTCAATCAGACGTGCAGGCCCAT	896
Qy	273	Leu	GluAsnTrpSerLysTrpArgGlySerValLeuSerArgArgLysArgHisTrpIle	292
Db	897	TT	GACTTATTTCGAGAATCGAGAGAAAATAAATTTGTCGACGCAAAAGGAATGATAAT	956
Qy	293	Ala	GlnIleThrAlaThrGluLeuAlaGlyThrThrValGlyLeuAlaTrpMetSer	312
Db	957	GCT	CAGTTACTCAAGACACTGCAATCAATGCAACTACTACAGACATGGTTTACATAGGC	1016
Qy	313	Thr	MetTrpSerPro---TyrSerValGlyValValGlnTrpHisSerTrpAsnLeuLeu	331
Db	1017	TCC	CTCTCTAGTCCGAGAAATCTGGCAGTTGTTCAGATCTATAGCAAAAGCACAGC	1076
Qy	332	Arg	ValAlaGlyThrMetAlaHisGluMetGlyHisAsnTrpGlyMetTrpHisTrpTrp	351
Db	1077	AT	GTGGCAANTACAAATGGCCCATCAGATGGGTATTAATCTGGGCATGAATGATGACAG	1136
Qy	352	Tyr	SerTrpLysTrp2toSerThrIleTrpValMetTrpLysAlaLeuSerTrpTyrrIle	371
Db	1137	GCT	TCTCTACTTGTGGTTCTAACAAATGCATTATG-----TCTACAAATATTATGAA	1190
Qy	372	Pro	ThrTrpTrpTrpSerSerTrpSerArgLeuSerTyrrTrpLysTrpTrpGluTrpLysLeu	391
Db	1191	TCT	CTT-----TCTGAGTTCCAGCTCTGTAGTCTCCAGGAACATCGGGAGTATCTCTT	1244
Qy	392	Ser	Asn-----TrpLeuTrpAsnAlaProLeuProThrTrpIleIleSerThrPro	408
Db	1245	AG	ACAGACACACCAATGCATCTCTCAACAACCTTCGGCGAAAGCTATTGTACACTTCCA	1304
Qy	409	Ile	TrpGlyAsnGlnLeuValGluMetGlyGluTrpTrpTrpTrpGlyThrSerGluGlu	428
Db	1305	GT	TGTGGAAATTAATTCTGTGGAGCGGGAGAAGAATGTGACTGTGCTCTCTCTGAGGAT	1364
Qy	429	Trp	ThrAsnIleTrpTrpTrpAlaLysThrTrpLysIleLysAlaThrTrpGlnTrpAla	448
Db	1365	TGT	CAAAATACCTGCTGATGCTGCAACTTGTAACTTGAACATGAGGCACAGTGTGAC	1424
Qy	449	Leu	GlyGluTrpTrpGluLysTrpGlnTrpLysLysAlaGlyMetValTrpArgProAla	468
Db	1425	TCT	GAGAGTCTGTGAGAAATGCAATTTAAGGAGCAGAGCAGCAATTCGCGGCAGCA	1484
Qy	469	Lys	TrpGluTrpTrpLeuProGluMetTrpAsnGlyLysSerGlyAsnTrpProTrpTrp	488
Db	1485	AAG	AATGACTGTGACTTCTTGAATCTGCACTGGCGGATCTGTAAGTGTCCAAAGCA	1544

RESIST 10

RESOL 10
US-08-514-014-1

03-08-214-014-1
: Sequence 1. Application VS/08514034

: Patent No. 5707829

; PAGE NO. 3107623
; GENERAL INFORMATION.

APPLICANT: Jacobs, Kenneth

APPLICANT: UACODS, REH
APPLICANT: McCov. John

APPLICANT: Kelleher, Kerry

APPLICANT: Carlin, McKeough

APPLICANT: CALTECH, MCNEOUGH.
TITLE OF INVENTION: DNA SEQUENCES AND SECRETED PROTEINS

/	TITLE OF INVENTION:	DNA SEQUENCES FOR
:	TITLE OF INVENTION:	ENCODED THERAPY

```

: TITLE OF INVENTION: EN
: NUMBER OF SEQUENCES: 1

```

1 NUMBER OF SEQUENCES: 1
: CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc. -- Legal Affairs

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02140

; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk

```

; COMPUTER: IBM PC compatible

```

;
; OPERATING SYSTEM:  PC-DOS/MS-DOS

```

```

; SOFTWARE: Patent In Release #1.0, Version #1.25

```

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/514,014

FILING DATE:

CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

NAME: Brown, Scott A

; REGISTRATION NUMBER: 32,724

; REFERENCE/DOCKET NUMBER: GI6000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8224

TELEFAX: (617) 876-5851

: INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2209 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

```
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 38..1447
US-08-514-014-1

Alignment Scores:
Pred. No.: 3,8e-44 Length: 2209
Score: 702.00 Matches: 177
Percent Similarity: 58.54% Conservative: 87
Best Local Similarity: 39.25% Mismatches: 161
Query Match: 14.90% Indels: 26
DB: 1 Gaps: 8

US-10-726-148A-15 (1-775) x US-08-514-014-1 (-2209)

QY 5 LeuLeuProValSer---LeuLeuLeuSerValAlaValSerAlaIleLysGluLeuPro 23
DB 89 CTGCTGCTGTACTTTGGCTCAATGTTCAAACTCAAGCAATAGCCATAAGCAACACCT 148
QY 24 GlyValLysLysGluValValTyrProIleArgLeuHisProLeuHisLysArgGlu 43
DB 149 GAATTAACGCTCCATGAATAGTTTCTCTAAATAAATCTTCAATTTTACAAAGAGAG 208
QY 44 AlaLysGluProGlu-----GlnGlnGluGlnTrpGluThrGluLeuLys 58
DB 209 ATCAAGAACACACACAGACAGAAAGCATGCGCAAGAGAGAAAGGTATGAACCTGAA 268
QY 59 TyrLysMetThrIleAsnGlyLysIleAlaValLeuTyrLeuLysLysAsnLysLeu 78
DB 269 TATCATGATCTTAAATGGAGAGAAATCATCTCTCCCTACAAATAAAGCAAGCACCTC 328
QY 79 LeuAlaProGlyTyrThrGluTrpTyrAsnSerThrGlyLysGluIleThrSer 98
DB 329 CTGGGCGCAGACTACATGAACATGTTACTCACCAGAGAGAGAGAAATTTACCAGAA 388
QY 99 ProGlnIleMetTyrTrpTyrTrpGlnGlyHisIleLeuAsnGluLysValSerTrp 118
DB 389 CCTGAGAACATGGAACACGTTACTATAAAGAGAAACATCTAAATGAAAGAAATTC 448
QY 119 AlaSerIleSerThrTyrArgGlyLeuArgGlyTyrTrpSerGlnGlyTrpGlnArgTyr 138
DB 449 GCCAGCAACAGTACTGTGACGGGTGAGAGATACCTTCACATCATCACCAGAGATAC 508
QY 139 TrpIleGluProLeuSerProIleHisArgTrpGlyGlnGluHisAlaLeuTrpLysTyr 158
DB 509 CAGATAAAACCTCTGAAAGACACAGCAG-----AAAGACATGCCCTCTTACATCT 562
QY 159 AsnProTyrGluLysAsn-----TyrTrpSerThrTrpGlyMet 171
DB 563 AACAGGAGGAGAACAGACCCAGCTAACACACATGTGGTGTGAAGACACTGACGGAA 622
QY 172 TrpGlyValLeuTrpAlaHisTrpLeuGlnGlnAsnIleAlaLeuProAlaThrLysLeu 191
DB 623 CAAGGCCCAAT-----CGAATCTCTAGATCACTCAAAAGCCAGAG-----664
QY 192 ValLysLeuLysTyrArgLysValGlnGluHisGluLysTyrIleGluTyrTrpLeuVal 211
DB 665 ---AAAGAGACTTTCTCGGCACAG-----AAATACATGATCTCTATTGGTG 712
QY 212 LeuTrpAsnGlyGluTrpLysArgTyrAsnGluAsnGlnTrpGluIleArgLysVal 231
DB 713 CTGGAATAATGCCCTTTATAAGAACTATAATGAGAATCTAACTCTGATAGAGCTTTG 772
QY 232 TrpGluMetAlaAsnTyrValAsnMetLeuTyrLysLysLeuAsnThrHisValAlaLeu 251
DB 773 TTTGATGATGAACCTACTCAATGTGATATATAACACCATAGATGTTCAAGTGCCCTG 832
QY 252 ValGlyMetGluIleTrpThrTrpLysTrpLysIleLysIleThrProAsnAlaSerTrp 271
DB 833 GTAGGTATGGAAATCTGTCTGATGGGATAGATAGAGGTGGTCCCGAGCAAGCACC 892
```

```
QY 272 ThrLeuGluAsnTrpSerLysTrpArgGlySerValLeuSerArgArgLysArgHisTrp 291
DB 893 ACGTTTGACAACTTCTCGAGATGGCACAGTCTTAACCTGGGG---AAAAAGATCCAGAC 949
QY 292 IleAlaGlnLeuIleThrAlaThrGluLeuAlaGlyThrThrValGlyLeuAlaTrpMet 311
DB 950 CATGCTCAGCTTCTCAGCGGGATTAGCTTCAACAATCGAGTGTGGGACTGGCAGCTTCA 1009
QY 312 SerThrMetTrpSerProTyrSerValGlyValValGlnTrpHisSerTrpAsnLeu 331
DB 1010 AATTCCTTGTTCCCTCCCTCTCGTGTCTGTTATTGAGGCTAAATAAGAAATTAATGTG 1069
QY 332 ArgValAlaGlyThrMetAlaHisGluMetGlyHisAsnTrpGlyMetTrpHisTrp 351
DB 1070 GCTCTTGTAGAGTGTATGTCACATGAGTGGCGCATGCTCTGTTATGCTGATGTTCCA 1129
QY 352 TyrSerTrpLysTrpProSerThrIleTrpValMetTrpLysAlaLeuSerTrpTyrIle 371
DB 1130 TTCAACACCAAGTGTCTCTGGCAGTGTGTGATGAATCAGTATCTGAGTTCAAAATTC 1189
QY 372 ProThrTrpTrpSerTrpSerArgLeuSerTyrTrpLysTrpTrpGluTrpLysLeu 391
DB 1190 CMAAGATTTCTAGTACTCTGCGTGCACATTTTGAAGATACCTTTTATCTCAGAAA 1249
QY 392 SerAsnTrpLeuTrpAsnAlaProLeuProThrTrpIleIleSerThrProIleTrpGly 411
DB 1250 CCAAGTGTCTGTCAGACACCTATTCTTACAAATATATGACACACCACTGTGTGG 1309
QY 412 AsnGlnLeuValGluMetGlyGluTrpTrpTrpTrpGlyThrSerGluGluTrpThrAsn 431
DB 1310 AACCCACTTCTAGAGTGGGAGAGACTGTATGTGGTCTCTCTAAGGAGTGTACCAAT 1369
QY 432 IleTrpTrpTrpAlaLysThrTrpLysIleLys 442
DB 1370 CTCTGTGTGAGCCCTTACGTTAACTGAAG 1402

RESULT 11
US-08-833-823-1
; Sequence 1, Application US/08833823
; Patent No. 5969093
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John
; APPLICANT: Kelleher, Kerry
; APPLICANT: Carlin, McKeough
; TITLE OF INVENTION: DNA SEQUENCES AND SECRETED PROTEINS
; TITLE OF INVENTION: ENCODED THEREBY
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc. -- Legal Affairs
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/833,823
; FILING DATE: 10-APR-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/514,014
; FILING DATE: 11-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: GI6000
; TELECOMMUNICATION INFORMATION:
```

QY 536 GluGlyGlySerLysTyrGlyTyrTrpArgValTrpTrpThrLeuLeuProTrpLys 555
Db 1637 GTCAGAGGGAATGATGTAAGCCACTCGAAGAAAGGAAATGGTCAAAAGATTCCATGTGCA 1696
QY 556 AlaAsnTrpThrMetTrpGlyLysLeuTrpTrpGlnGlySerTrpAsnLeuProTrp 575
Db 1697 GCAAGAGTAAAGTGTGGCAGGTATATCTGCACAGAGAGAC-----1741
QY 576 LysGlyArgIleValThrTrpLeuThrTrpLysThrTrpProGlnTrpThrSerGln 595
Db 1742 -----ACAATGTCATCCGATCCCA-----CTGGACCCA 1772
QY 596 GlnLeuGlyMetValAlaAsnGlyThrLysTrpGlyTrpAsnLysValTrpIleAsnAla 615
Db 1772 GATGGTGAATGCTGAACCTCGAACAATAATGTGGAGATGAATGTGTGCACCAACGGT 1832
QY 616 GluTrpValTrpIleGlnLysAlaTyr---LysSerThrAsnTrpSer 630
Db 1832 CAGTGTGTAATGTGCAGACAGCCCTACTGATCAAGCACTGGCTCTCT 1879

RESULT 15

US-09-026-001A-7
; Sequence 7, Application US/09026001A
; Patent No. 6413760
; GENERAL INFORMATION:
; APPLICANT: Boodhoo, Anechand
; APPLICANT: Seehra, Jasbir
; APPLICANT: Shaw, Gray
; APPLICANT: Sako, Dianne
; TITLE OF INVENTION: HIGHLY PURIFIED MOCARHAGIN, A COBRA VENOM
; TITLE OF INVENTION: THERAPEUTIC USES THEREOF
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/026,001A
; FILING DATE: 18-FEB-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G152938
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2297 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-026-001A-7

Alignment Scores:
Pred. No.: 8,26p-42 Length: 2297
Score: 671.50 Matches: 205
Percent Similarity: 47.89% Conservative: 101
Best Local Similarity: 32.08% Mismatches: 284
Query Match: 14.25% Indels: 49
DB: 4 Gaps: 12

US-10-726-148A-15 (1-775) x US-09-026-001A-7 (1-2297)

QY 1 MetLeuGlnGlyLeuLeuProValSerLeuLeuLeuSerValAlaVal-----16
Db 85 ATGATCCAGACTCTCTTG-----GTAATATATATGCTTAGCGGTTTTTCCATATCAA 135
QY 17 -----SerAlaIleLysGluLeuProGlyValLysLysTyrGluValValTyrProIle 34
Db 136 GGGAGCTCTATATCTGGAATCTGGAAATGTTAATGATTATGAGTTGTTGATCCACAA 195
QY 35 ArgLeuHisProLeuHisLysArgGluAlaLysGluProGlnGlnGlnGlnTrpGlu 54
Db 196 AAGAGTCCCTGCATCTGCTCAAGAGGAGGAGTTCAGAACTCTCAGCAGAGACCAAGTATGAA 255
QY 55 ThrGluLeuLysTyrLysMetThrIleAsnGlyLysIleAlaValLeuLysLeuLysLys 74
Db 256 GATACAAATGCATATGAAATTCAGATGAATGGAGAGCCAGTAGTCTCTTACATTAGAAAGA 325
QY 75 AsnLysAsnLeuLeuAlaProGlyTyrThrGluThrTyrTyrAsnSerThrGlyLysGlu 94
Db 316 AATAAAGGACTTTTTCAGAGAGATTACACTGAACCTCATTTGCCCTGATGCGCAGGAA 375
QY 95 IleThrThrSerProGlnIleMetTrpTrpTrpTyrGlnGlnGlyHisIleLeuAsnGlu 114
Db 376 ATTACAAACAGCCCTCCGGTTCAGGATCACTGCTATTATCATGTTACATTACAGATGAA 435
QY 115 LysValSerTrpAlaSerIleSerThrTrpArgGlyLeuArgGlyTyrTrpSerGlnGly 134
Db 436 GCTGACTCAAGTCAATCATCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTC 495
QY 135 TrpGlnArgTyrTrpIleGluProLeuSerProIleHisArgTrpGlyGlnGluHisAla 154
Db 496 GGGAGACATACATTTATTGAGCCCTTGAGCTTTTCGAC-----AGTGAATCTCAGCA 549
QY 155 LeuTrpLysTyrAsnProTrpGluLysAsnTyrTrpSerThrTrpGlyMetTrpGlyVal 174
Db 550 ATCTACAAAGATGAAATATGAGAA---AACGAGGATGAGACCCCGCAAACTGTGGGTA 606
QY 175 LeuTrpAlaHisTrpLeuGlnGlnAsnIleAlaLeuProAlaThrLysLeuValLysLeu 194
Db 607 ACGGAGACTACTTGGGAGTCAGATGATCCATGCAAGAAAGACCTCTCAGTTAATAACACT 666
QY 195 LysTrpArgLys---ValGlnGlnHisGluLysTyrIleGluTyrTrpLeuValLeuTrp 213
Db 667 CCTGAACAGACGCGGTACTTGCAGGCCCAAAATACATCGAGTTTACGSGTGTGGAC 726
QY 214 AsnGlyGluTrpLysArgTyrAsnGluAsnGlnTrpGluIleArgLysArgValTrpGlu 233
Db 727 AACAGAAATGTACAGGTATTACAAACGCAATGAACCTGCTATATAAAGAGAGTATATGAA 786
QY 234 MetAlaAsnTyrValAsnMetLeuTyrLysLysLeuAsnThrHisValAlaLeuValGly 253
Db 787 ATGCTCAACGCTGTAAATACGAGTACAGACCTTTGAAATTCACATAACACTGATGGC 846
QY 254 MetGluIleTrpThrTrpLysTyrLysIleLysIleThrProAsnAlaSerTrpThrLeu 273
Db 847 CTAGAAATTTGGTCCCAACCATGATAAGTTTGAAGTGAAGCCAGTAGCGGTGCCCTTTG 906
QY 274 GluAsnTrpSerLysTrpArgGlySerValLeuSerArgArgLysArgHisTrpIleAla 293
Db 907 AAATCATTTTCGAGATTGGAGAGAAACAGTTTGTGCCACGCAAAAGGAATGATACGCT 966
QY 294 GlnLeuIleThrAlaThrGluLeuAlaGlyThrValGlyLeuAlaTrpMetSerThr 313
Db 967 CAGTTACTCAGCGGCAATGACTTCAATGGACCTGTGTGGGAATGCTTACAGGGCACCC 1026
QY 314 MetTrpSerProTyrSerValGlyValValGlnTrpHisSerTrpAsnLeuLeuArgVal 333
Db 1027 CTCTGCACCTCAGAAATCTGTAGCAGTCTTTCAGGATTATACCGGAAAAAATAGCATGGTG 1086
QY 334 AlaGlyThrMetAlaHisGluMetGlyHisAsnTrpGlyMetTrpHisTrpTrpTrpSer 353
Db 1087 GCATCTACATGGCCCATGAGTTGGTTCATATCTGGGCCCTTCATCATGACGAGCTTCC 1146

QY 354 TrpLysTrpProSerThrIleTrpValMetTrpLysAlaLeuSerTrpTyrIleProThr 373
 Db 1147 TGTATTTCAGTCTTAGACCATGCTATGCTAAGGGAGGACT-----GCACCTGCC 1200
 QY 374 Trp---TrpSerSerTrpSerArgLeuSerTrpLysTrpGluTrpLysLeuSer 392
 Db 1201 ITTCAGTTTCAGTCTTGTAGTCCGGAGTATCGGGAGTATCTTCTTAGAGAAAGACCA 1260
 QY 393 AsnTrpLeuTrpAsnAlaProLeuProThrTrpIleIleSerThrProIleTrpGlyAsn 412
 Db 1261 CAATGCATCTCTCAACAAACCCCTTGACACAGATACCTGTTCTCCCAATTTGTGGAAT 1320
 QY 413 GlnLeuValShuMetGlyGluTrpTrpTrpGlyThrSerGluGluTrpThrAsnIle 432
 Db 1321 TACTTTGTGGAGGAGGAGAGAAATCTGACTGTGGCTCTCTCGCGATGTCAAAGTGCC 1380
 QY 433 TrpTrpTrpAlaLysThrTrpLysIleLysAlaThrTrpGlnTrpAlaLeuGlyGluTrp 452
 Db 1381 TGCTCGATGCTGCAACTGTGTTAGTTAAG-----GGAGAAGAA 1419
 QY 453 TrpGluLysTrpGlnTrpLysLysAlaGlyMetValTrpArgProAlaLysTrpGluTrp 472
 Db 1420 GCAGAA-----TGCCGGGCGACAAAGGATGACTGT 1449
 QY 473 TrpLeuProGluMetTrpAsnGlyLysSerGlyAsnTrpProTrpTrpArgTrpGlnVal 492
 Db 1450 GACTTGCCTGAACCTGCACTGGCGACTCTGTGGAGTGTCCCGACGACAGCTTGCGAG 1509
 QY 493 AsnGlyTrpProTrpHisGlyLysGlyHisTrpLeuMetGlyThrTrpProThrLeu 512
 Db 1510 AATGACATCCATGTCAAAACAAACAAAGGTTACTGCTACAATGGGGCATGTCCACCTTC 1569
 QY 513 GlnGluGlnTrpThrGluLeuTrpGlyProGlyThrGluValAlaTrpLysSerTrpTyr 532
 Db 1570 ACAACCAATGTATTGCTCTCATGGGACAGATTTACTGTGATGCCAGATGATGTCT 1629
 QY 533 AsnArgAsnGluGlySerLysTyrGlyTrpArgArgValTrpTrpThrLeuIle 552
 Db 1630 GACTTGAACGTGAGAGGGAATCATGTAAAGCCACTGCAGAAAGGAAAATGGTGCAAGAT 1689
 QY 553 ProTrpLysAlaAsnTrpThrMetTrpGlyLysLeuTrpTrpGlnGlyGlySerTrpAsn 572
 Db 1690 CCATGTGACGCAAGGATGTAAAGTGTGGCAGGTTATATCTGCACACAGAGAGAC----- 1743
 QY 573 LeuProTrpLysGlyArgIleValThrTrpLeuThrTrpLysThrTrpProGluTrp 592
 Db 1744 -----ACAAATGTCATGCCGATTCCCA----- 1764
 QY 593 ThrSerGlnGluIleGlyMetValAlaAsnGlyThrLysTrpGlyTrpAsnLysValTrp 612
 Db 1765 CTGGACCCAGATGGTGTAAAGCTGAACCTGGACAAATAATGGAGATGGAATGGTGTGC 1824
 QY 613 IleAsnAlaGluTrpValTrpIleGluLysAlaTyr---LysSerThrAsnTrpSer 630
 Db 1825 AGCAACGGTCACTGTGTTAATGTGCAGACAGCTACTGATCAAGCACTGGCTTCTCT 1881

Search completed: June 10, 2004, 01:00:55
 Job time : 189 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2004, 17:58:05 ; Search time 21 Seconds
(without alignments)
3549.925 Million cell updates/sec

Title: US-10-726-148a-15

Perfect score: 4712

Sequence: 1 MLQGLLPVSLLSVAVSAIK.....PPTVKNKPNMSTPKWNPKA 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 3

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	773.5	16.4	776	2	S28258
2	755.5	16.0	789	2	S28259
3	695.5	14.8	617	2	S48160
4	693.5	14.7	609	2	S55270
5	686.5	14.6	826	2	A60385
6	671	14.2	571	2	S24789
7	654	13.9	610	2	JC7530
8	633	13.4	616	2	A55796
9	552.5	11.7	549	2	S48169
10	551	11.7	407	2	S68260
11	543	11.5	411	1	HYNSFA
12	543	11.5	414	1	HYESAC
13	540	11.5	414	2	S41609
14	536	11.4	478	2	JC4880
15	535	11.4	481	2	S43125
16	534.5	11.3	478	2	A43296
17	532	11.3	414	2	S41608
18	526	11.2	481	2	JC4342
19	516	11.0	480	1	A30065
20	484	10.3	478	2	JQ1301
21	475	10.1	419	2	A59414
22	468	9.9	419	2	S41607
23	457.5	9.7	903	2	S60257
24	436.5	9.3	416	2	A37877
25	401.5	8.5	814	2	G02390
26	395	8.4	655	2	JC7850
27	367.5	7.8	429	2	A42972
28	303.5	6.4	825	2	S55060
29	300	6.4	1146	2	T02766

ALIGNMENTS

RESULT 1

S28258

androgen-regulated epididymal protein precursor - crab-eating macaque

C:Species: Macaca fascicularis (crab-eating macaque)

C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 21-Jul-2000

C:Accession: S28258

R:Perry, A.C.F.; Jones, R.; Barker, P.J.; Hall, L.

Biochem. J. 286, 671-675, 1992

A:Title: A mammalian epididymal protein with remarkable sequence similarity to snake ve

A:Reference number: S28258; MUID:93038491; PMID:1417724

A:Accession: S28258

A>Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-776 <PER>

A:Cross-references: EMBL:X66139; NID:q38062; PIDN:CAA46929.1; PID:q38063

C:Superfamily: mouse meltrin alpha; disintegrin homology

F:1-25/Domain: signal sequence #status predicted <SIG>

F:26-776/Product: androgen-regulated epididymal protein #status predicted <MAT>

F:336/Region: defective catalytic site

F:401-483/Domain: disintegrin homology <DIS>

Query Match	16.4%	Score	773.5	DB	2	Length	776
Best Local Similarity	30.8%	Pred. No.	1.4e-48				
Matches	247	Conservative	137	Mismatches	322	Indels	97
Gaps	20						
QY	1	MLQGLLPVSLLSVAVSAIKELPGVKYEVVYPIRLHPLHKRE---	AKPEQCEQWETEL	57			
DB	1	MLPGCIFLMILLQLQVKE-KVILGVGGQLVVPKPLPKQKRDI	GHTHDDDIETYEEL	59			
QY	58	KYNTINGKIAVYLYKKNKLLAPGTETYYNSTGKEITTSPOIMW	WYVYQGHILNEKVS	117			
DB	60	MYEIKLNKRLTVLHLRSEFLGSNYSETFYSMKGEAFTRHLQIM	DHCFYQGSIVHEYDS	119			
QY	118	WASISTWGLRGVMSQGWRYMTEPLSPHRCQEHALMKYN---	PWEKNYSTWGWGV	174			
DB	120	AAISITCNGLGFRRVNDQRIEIPVK--YSDEGHLVFKYNPRV	PVYVANYSC	171			
QY	175	LWAHLQQMIAL-----PATKLVLKRWKQVSHKQHEKIVYLV	LWNGEKRYNE	222			
DB	172	-----ELNFTKTPVGDTESESGDPKMKAI-----	HNEKYIELFIVADTVYRN	217			
QY	223	NOWETKRWEMANYVMYKLNTHVALVGMIEWTWKIKITPNASWT	LENWSKWRGS	282			
DB	218	PHNKLNRNIGWVFNVMYKTLNHLVLTGTEIWHEDKIELHSN	ETTLRPFSSQWER	277			
QY	283	VLSREKREHIAQLITATELAGTTVGLAWKSTWSP--YSVGV	VQWHSWNLIR	337			
DB	278	ILKTRDFHVVLLSGKWIYTHVQGISYPAGMCLPYSTSIK---	DLPTNTIANRM	333			
QY	338	AHEMGNHGMHMYWYWKWNPSTIWMKALSYIPTW-WSSWSR	LSYKWKWKLNLW	396			
DB	334	AHOLGHNLMQHDPEFCTCPGKCVMSDGS--IPALNFKSC	QNYHYQLKDYKTCML	391			


```

Qy 178 HWLQONTALPATKLVKLVKRWKVQBEHEKYIEYVLVLWNGEWKRYENQWBEIRKRVWEMANY 2317
Db 176 NWESDELK-KASQLVATS-EQORSYKKYIEFVVVADYIMRYKYNNDSTAVRRRIIVEINV 2333
Qy 238 VNMYLKLINTHALVGMIEITWKKIKITPPNASWTLENNWSKRGSVLSRKHMTLAQILT 297
Db 234 LNMVITVNIHVALTHLISWSTROQITVQSAADVTLDLPGMRANKLLTRKXKHDNAQLPT 293
Qy 298 ATELAGTITVGLAWMSWTMSP-YSVGVGWQHSNNLLRVLVAGTMAHEHGMWGMWVWYSWKW 356
Db 294 GINLNGQTLGIARMGCMCSPNSGVIGDYCKNYLLVALITMAHELGNLGMDDHNGCNC 353
Qy 357 PSTITWMMKALSWIPTW-WSSWSRLSVTKWWEKLSNNLWNAPLPTWITISTPIWGNQLV 415
Db 354 PDTSCIM-SAVAGPEPEVFSCNSRNDYRSPNSQSKCIDNKPLKTDIVSPSVCGNYEV 412
Qy 416 EMGEWMMWMTGSEETNIIWWAKTKIKATWQALGEWEKQWKAGWVRPAKWEWMLP 475
Db 413 EVGECDGSRITYCENPCNATTKLITGSCQADGECNOCQPRPARTECRKIDDCDVP 472
Qy 476 EMWNGKSNWPWQVNGVPMWHGKHGLMGTPTPLQEQWTELWGP-----GTEVANKSW 531
Db 473 EYCTGQSGECLDVFQRNGQPYQSNNGCYCNGNCPTLKNQCIHLWKPAPGAVGVAPDVC 532
Qy 532 YNRNEGGS 539
Db 533 FEDNQNGT 540

RESULT 4
S55270
C:catrocollastatin precursor - western diamondback rattlesnake
C:Species: Crotalus atrox (western diamondback rattlesnake)
C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 21-Jul-2000
C:Accession: S55270
R:Zhou, Q.; Smith, J.B.; Grossman, M.H.
Biochem. J. 307, 411-417, 1995
A:Title: Molecular cloning and expression of catrocollastatin, a snake-venom
A:Reference number: S55264; MUID:95251603; PMID:7733877
A:Accession: S55270
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-609 <ZHO>
A:Cross-references: GB:U210033; NID:g710353; PIDN:AAC59672.1; PID:g710354
C:Superfamily: mouse melittin alpha; disintegrin homology

```

Db 344 PGVCYCEPREGGGCGIMTESGSKFPR-FSRCSKIDLESFVTKPQTGCLTQVFNRFVGG 403
QY 408 PINGNOLVGEVGEWWWGTSSEETINWNAKTKI.KATQWALGEVWEKQWKKAGVWRP 467
Db 404 PFCGNLFVEHGEQDCGTPOCCQPCCNATTCQDVKGABCASGTCCECKVKPAGEVCL 463
QY 468 AKWEWMLPEWNGKSGNWPWVRWQVNGWPHHKGHWLMTPTLOQWTELWGPGEVA 527
Db 464 SKDKCOLLEECGRKPTCPEDAFQNGTTPC--PGYCFDGSCTPLAQCDRLWFGARVA 521
QY 528 WKSWYNNRNEGSGYGVWRRVWMLIPKANKTW-----GKLWQGGSNLIPKGRIV 580
Db 522 ADSCTYFS-----IPPCNRMYSGRINRCGALYCBGG--QKPLERSFC 563
QY 581 TMLTWTMTWPEWTSQETG-----MWANGTKWGNKWNAEWNAEWIEKAYESTWSS 631
Db 564 TFS-----NHGVALGTGTSNDTFFELVIGTKCEGKVCMDGSCQDL-RVYSENCSA 617
QY 632 KXKGHAVWHHELOQWBEHIPPWW-----WSSVYVWHSYVGVLPNAVIVV 681
Db 618 KCNNHGVCHKRECHCKHGWAPPNCVQRLADVSDQAASLSPVSVVVLVILVAAVIV 677
QY 682 VAMVIRHQSREKQKWRPLSTTG-TPEHKQKRPQWVAVQPEMSQKPHVWLPVE 740
Db 678 AGVITIRKAPRQIRRSVAPKPSGLSNP-FYTRDSSLPAKNRPDPSE-----TVS 729
QY 741 GNEP--PASWHKWTNALPPTVWKNWMS.TPKWSNPK 774
Db 730 TNQPPRIAPKRPAPPAGVSSPLVPVYA-PK 764
RESULT 6
S24789
jararagin C precursor - jararaca (fragment)
N;Alternate names: single chain botrocetin
N;Contains: disintegrin-like 28K protein; hemorrhagic proteinase (EC 3.4.24.-)
C;Species: Bothrops jararaca (jararaca)
C;Date: 20-Feb-1995 #sequence revision 29-Aug-1997 #text_change 09-Jun-2000
C;Accession: S24789; J02245; A44463; A37958; J02373
R;Paine, M.J.
submitted to the EMBL Data Library, August 1992
A;Reference number: S24789
A;Accession: S24789
A;Molecule type: mRNA
A;Residues: 1-571 <PA1>
A;Cross-references: EMBL:X68251; NID:962467; PID:G62468
R;Usami, Y.; Fujimura, Y.; Miura, S.; Shima, H.; Yoshida, E.; Yoshioka, A.; Hirano, K.;
Biochem. Biophys. Res. Commun. 201, 331-339, 1994
A;Title: A 28 kDa-protein with disintegrin-like structure (jararagin-C) purified from B
A;Reference number: J02245; MUID:9425699; PMID:8198592
A;Accession: J02245
A;Molecule type: protein
A;Residues: 360-571 <USA>
A;Experimental source: venom
R;Paine, M.J.; Desmond, H.P.; Theakston, R.D.; Crampton, J.M.
J. Biol. Chem. 267, 22869-22876, 1992
A;Title: Purification, cloning, and molecular characterization of a high molecular weight
ily.
A;Reference number: A44463; MUID:93054601; PMID:1385408
A;Accession: A44463
A;Molecule type: mRNA
A;Residues: 1-23; 'Q', 25-92; 'G', 94-131; 'G', 133-169; 'Q', 171-571 <PA2>
A;Cross-references: GB:X68251
A;Experimental source: venom gland
A;Note: sequence inconsistent with the nucleotide translation
R;Fujimura, Y.; Titani, K.; Usami, Y.; Suzuki, M.; Oyama, R.; Matsui, T.; Fukui, H.; Sug
Biochemistry 30, 1957-1964, 1991
A;Title: Isolation and chemical characterization of two structurally and functionally di
A;Reference number: A37958; MUID:91129280; PMID:1993206
A;Accession: A37958
A;Molecule type: protein

A;Residues: 360-372, 'E', 374-378, 'X', 380-384 <FUG>
A;Note: 361-Val was also found
C;Comment: Inhibits collagen- and ADP-induced platelet aggregation.
C;Superfamily: mouse meltrin alpha; disintegrin homology
C;Keywords: hydrolysis; metalloproteinase; venom; zinc
F;360-571/Product: jararagin C #status experimental <MAT>
F;362-444/Domain: disintegrin homology <DIS>
F;295,299,305/Binding site: zinc (His) #status predicted
F;296/Active site: Glu #status predicted

Query Match 14.2%; Score 671; DB 2; Length 571;
Best local Similarity 33.6%; Pred. No. 2.9e-41;
Matches 195; Conservative 102; Mismatches 260; Indels 24; Gaps 9;

QY 50 QEQWETELKYKMTINGKIALVILAKNNKLLAPGYTETVYNSTGKEITTSPOIMWVYCG 109
Db 9 QPKYEDAWQVEFKVNGEPVVLHLEKNKGLFSKSEIHSPDGREITTYPPVDEHCYHG 68
QY 110 HILNKYSWASISITWRGLRGYSQWQRYETIEPLSPIHRMQEHALMKYNWKNYSTW 169
Db 69 RIENDADSTASISACNGLKGYFKLQRETYFIEPLKLPD--SEAHAVFYENVEKDEAP- 125
QY 170 GMGVLWAHMLQONIALPATKLVKRWKVOEHKXYEYLYLVNKGWKNYENQWEIRK 229
Db 126 KMGCVT-QNWKSEPIKASQLAFTABQORYDPKYIEFFVVDQGTVKNNGDLDTKA 184
QY 230 RVNEMANYVMLYKKNLTHVALVGMELTWKWKIKITPNASMTLENNKSWGSLSRKR 289
Db 185 RMELANIVNEIFRYLYMHVALGLEIWSNGDKITVDPDVTYLSFAEMKRTDLLTRK 244
QY 290 HWIAQLITATELAGTTVGLAWMTMSP--YSGVYVQWHSWNLRLVAGTMAHEMGNWGM 348
Db 245 HDNAQLLTALDFNGPTIGYVIGSMCHPKRSVGIVQDYSPLNVAVVIMAHMGHNLGIH 304
QY 349 HWYWSKWPSTIWMVXALSWYIPTWWSWSRLSYKWNWKLNS---WLNAALPTWII 405
Db 305 HDTGSCSGYPCYIMGPTIS-NEPS--KFFSNCSYIQWDFIMHNPECIINEPLGTDII 361
QY 406 STPIWKNOLVEMGEWMMWGTSEETNIWNAKTKIKATQWALGEVWEKQWKKAGVW 465
Db 362 SPPVCNLELLEVBECDCGTPENCCECDATCKLSGSCQCGHGDCECQCKFSKGTEC 421
QY 466 RPAKWEWMLPEWNGKSGNWPWVRWQVNGWPHHKGHWLMTPTLOQWTELWGPGE 525
Db 422 RASMSCECPAECHCTGQSCEPADVFEKNGQPCLDNYCYNGNCPIMVHCYALFGADVY 481
QY 526 VANKSWYRNREGSKYGVWRRVWMLIPKANKTWKGLWQGS--WNLPWKGRIVTWL 583
Db 482 EAEDSCFKDNQKNYGYCYRKBNGKKI.PCAPEEYKCGRLYCKNSPQNNPCK----- 534
QY 584 TWKTMPEWTSQETIGMVANGTKWGNKWVNAEWVWIEKAY 624
Db 535 ----MFYSNDEHKGWVLPCTKADGVKCSNGHCVDTATAY 571

RESULT 7
JC7530

vascular apoptosis-inducing protein 1 - western diamondback rattlesnake
C;Species: Crotalus atrox (western diamondback rattlesnake)
C;Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 01-Mar-2002
C;Accession: JC7530; PC7105
R;Masuda, S.; Ohta, T.; Kaji, K.; Fox, J.W.; Hayashi, H.; Araki, S.
Biochem. Biophys. Res. Commun. 278, 197-204, 2000
A;Title: cDNA cloning and characterization of vascular apoptosis-inducing protein 1.
A;Reference number: JC7530; MUID:20525424; PMID:11071872
A;Contents: Crude venom gland
A;Accession: JC7530
A;Molecule type: mRNA
A;Residues: 1-610 <MAS>
A;Cross-references: DDBJ:AB042840
A;Accession: PC7105
A;Molecule type: protein
A;Residues: 240-251;274-283;378-385;507-519 <MA2>

C;Comment: This protein, as a homodimeric protein, belongs to snake venom metalloprotease
a specific substrate, induces apoptosis in vascular endothelial cells.

C;Genetics:

A;Gene: vapi

C;Superfamily: mouse meltrin alpha; disintegrin homology

C;Keywords: apoptosis

Query Match 13.9%; Score 654; DB 2; Length 610;
Best Local Similarity 32.3%; Pred. No. 5.5e-40;
Matches 204; Conservative 104; Mismatches 274; Indels 70; Gaps 13;

QY 1 MQQLLPVSLLSAV-----SAIKELPGVKYEVVPIRLHPLHREAKEPEQOOWE 54

DB 1 MQVLL---VTISLAVFPYQGSVILESGNVDYEVYRKVTALPKGAV-----QPKYE 52

QY 55 TELKYMTINGKIAVLVLYKKNNLLAPGYTETYNSTGKEITTSPOIMWVYQGHINE 114

DB 53 DAMQVEFKNGEPVVLHLEKNGKLFSEDSYSETHSPDGREITTPVEDHCYHGRLEND 112

QY 115 KVSASISTWGLRGYMQGRYWIPLSPIHRWGQHALKYNPKYKYSTWGMGV 174

DB 113 ADSTASISACNLGKHFKLQEMYLEPLKLPD--SEAHAVKYENVEKEDAP-KMCGV 169

QY 175 L--WAHM-----LQCNIALPATKLVKLKWRKVQEHKEIYVLYLNGEKRYNENQWE 226

DB 170 TONWESYEPKKAQSNLTPEQORVYNAK-----KYVKLPVADYIMYLKYGRNLTA 221

QY 227 IRKXVWEMANYVMYLKXNLTHVALVGMETWKKIKITPNASWTLNWSKWSVLSR 286

DB 222 VRTMYDIVNVITPIYHRMNFVALVGLIWSNTDKIIVQSSADVTLDLFAKWRATDLS 281

QY 287 RKRWIAQLITATELAGTTVGLAWN---STWSPSYSGVQVQHSWNLRLVAGTMAHEMG 342

DB 282 RKSDNAQLLTGINFNGTAGLHGGICNTM---YSAGIVQDHSKIHLVAIAWAHEMG 338

QY 343 HNWGMWVWYKSWKWPSTIWNVWKAISWY-PTWWSWSLSLYKWKWVWKLNLWNAPEPT 402

DB 339 HNLGMDHDKDTCTGTRPCVMAGALSCEASFLPDCSQKHREPLIKNMPQILKPKPT 398

QY 403 WIITPTLGNOLVNGEHWWTGTSEWTNIWNAKTWIKATWQALGEWKEWQWKAG 462

DB 399 DVSPAVCGNYFVSVGECDGSPRTCRDPCCDATTCKLRQAQCAEGLCDQCRCFKAG 458

QY 463 MVWRPAKWWMLPBMWAKSGNMPWVQVNGVWPHHSGHMLXGTWPTLQEOITELWGP 522

DB 459 TECAADECDMAVCTGRSAECT-DRFQNGQCKXNNGYCNGKCEPMADQCIALFGP 517

QY 523 STEVAKSWYNNRNGSKYGYRRVWMTLIPWKNWMTWGLWQGGSWNLPLWKGRIVTW 582

DB 518 GATVSDAQCFQFNRGNHYGCRKEQNTKIACEPDQVCKGRLYCFPNS-----565

QY 583 LTWKTWKE-----WTSQEIGNVANGTKWGNKYVINAENWVIEKAY 624

DB 566 -----PENKNPKNIYSPNDEDKGVLPFGTKCADRKACNSGQCVDTTPY 610

RESULT 8

A55796

ecarin precursor - saw-scaled viper

C;Species: Echin carinatus (saw-scaled viper)

C;Date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 21-Jul-2000

C;Accession: A55796

R;Nishida, S.; Fujita, T.; Kohno, N.; Atoda, H.; Morita, T.; Takeya, H.; Kido, I.; Paine

Biochemistry 34, 1771-1778, 1995

A;Title: cDNA cloning and deduced amino acid sequence of prothrombin activator (ecarin)

A;Reference number: A55796; MUID:95151760; PMID:7849037

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-616 <NIS>

A;Cross-references: GB:D32212; NID:G717090; PIN:BA06910.1; PID:G717091

C;Superfamily: mouse meltrin alpha; disintegrin homology

F;404-486/Domain: disintegrin homology <DIS>

F;338/Active site: Glu #status predicted

Query Match 13.4%; Score 633; DB 2; Length 616;

Best Local Similarity 32.2%; Pred. No. 1.9e-38;

Matches 206; Conservative 98; Mismatches 297; Indels 38; Gaps 12;

QY 1 MQQLLPVSLLSV-----AVSAIKELPGVKYEVVPIRLHPLHREAKEPEQOOWETE 56

DB 1 MQVLL-VIICLAVFPYQGSIIILASGNVDYEVVYQKVTALPKGAVQOPEQ--KYEDA 57

QY 57 LKVMKTINGKIAVLVLYKKNNLLAPGYTETYNSTGKEITTSPOIMWVYQGHINEKY 116

DB 58 MOYEFVKGEPVVLHLEKNGKLFSEDSYSETHSSDDREITTPSVBDHCYHGRIONDAE 117

QY 117 SWASISTWGLRGYMQGRYWIPLSPIHRWGQHALKYNPKYKYSTWGMGV 176

DB 118 STASISACNLGKHFKLQETVIEPLKIPD--SEAHAVKYENIE-NEDEAPKMGVTVQ 174

QY 177 AHWLQONLALPATKLVKLKWRKVQEHK-----KYEYLYLNGEKRYNENQWEIRKRV 231

DB 175 DNWESDE---PIKKTIGL---IVPHERKFEKKFIELVVDHSMVTKYNDSTAIRTWI 228

QY 232 WEMANYVMYLKXNLTHVALVGMETWKKIKITPNASWTLNWSKWSVLSRBRHW 291

DB 229 YEMLVNVEIYLPFNRVALVGLFVWCGDLINVTSADTLHSGFGEWRASDLNKRKRD 288

QY 292 IAQLITATELAGTTVGLAWMSTMW-SPYSGVQVQVQHSWNLRLVAGTMAHEMGHNMHW 350

DB 283 HAQLLTNTLDHSTIGITFVYCMCKSDRSVELILDYSNITFNWYIAHEMGHSLGMLHD 348

QY 351 WYSWKPSTIWNVWKAISWYPTWWSWSLSLYKWKWVWKLNLWNAPEPTWISTPTW 410

DB 349 TKFCTCGAKPCIMFGEKESIPPPKEFSSCSYDQNKYLLKYNPKCILDPRLRXDIASEAVC 408

QY 411 GNQLVEMGEWMMWSTSEWTNIWNAKTWIKATWQALGEWKEWQWKAGWVRPAKW 470

DB 409 GNEIWEGBECDCGSPADCRNPPCCDATTCKLRQAQCAEGLCDQCRCFKAGTECRPAD 468

QY 471 EWLPEPMWNGKSGNMPWVQVNGVWPHHSGHMLXGTWPTLQEOITELWGPTEVAVKS 530

DB 469 DCDVAEHCTGQSAECFRNEFORNGQPCLNNSGYCNGDPCIMLNQCIALLFSFSAITVADS 528

QY 531 WYNNEGSKYGYRR-----VWMTLIPWKNWMTWGLWQGGSW--NLPWKGRIVTW 585

DB 529 CFQNLQSGYGYCTKEITGYGKRFPCAPQDVCKGRLYCLDINSFKKNRCKNDY-----582

QY 586 KTWPEWTSQSIGVANGTKWGNKYVINAENWVIEKAY 624

DB 583 -----STADENKGIVPGTKCEDGKVCINRKCVDVNTAY 616

RESULT 9

S48169

metalloproteinase (EC 3.4.24.-) H-II precursor - carpet viper (fragment)

N;Contains: disintegrin

C;Species: Echis pyramidum leakeyi

C;Date: 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 09-Jun-2000

C;Accession: S48169

R;Paine, M.J.I.; Moura-Da-Silva, A.M.; Theakston, R.D.G.; Crampton, J.M.

A;Title: Cloning of metalloproteinase genes in the carpet viper (Echis pyramidum leakeyi)

A;Reference number: S48169; MUID:95010025; PMID:7925363

A;Molecule type: mRNA

A;Residues: 1-549 <PAI>

A;Cross-references: EMBL:X78971; NID:G763094; PIDN:CAA55566.1; PID:G763095

C;Superfamily: mouse meltrin alpha; disintegrin homology

C;Keywords: hydrolase; metalloproteinase; venom; zinc

F;1-138/Domain: propeptide (fragment) #status predicted <PRO>

F;139-549/Product: metalloproteinase H-II #status predicted <PRO>

F;350-432/Domain: disintegrin homology <DIS>

F;284/Active site: Glu #status predicted

Query Match 11.5%; Score 543; DB 2; Length 414;
Best Local Similarity 35.4%; Pred. No. 7e-32;
Matches 151; Conservative 82; Mismatches 173; Indels 20; Gaps 9;

Qy 1 MLOGLLPVSLLSV-----AVSAIKELPGVKYKVVYPIRLHPLHKKRAKEPEQOEOWETE 56
 Db 1 MIEVVL-VTICLAVFPYQGSIILESQVNDYEVVYPRKVTALPKGAV-----QPKYEDA 54

Qy 57 LKYMRTINGIAVLVLYLKKNNKLLAPGYTETYYNSTGKEITTSPOIMWWWYQGHILNEKV 116
 Db 55 MOYEFVNGEPVVLHLEKKNELFSKDYSETHYSPDGREITINPAVEDHCYHGRIONDAD 114

Qy 117 SWASISTWRGLRGYSQGWQRYWIEPLSPIHRWQOEHALWKYNPWEKNYKYSTWGMWGLW 176
 Db 115 STASISACGLKGHFYKLGQEDYILEPLESD--SEAHAVFLENVEKDEAP--KMGCVT- 170

Qy 177 AHMLQQNIAPATKLVKLKWRKQOE--EKYIEYLVLMNGENKRYNENOWEIRKRVWEM 234
 Db 171 QNWESYB--PIKKASDLNLKEDQNLPPQRYIELVVADVADHRVFKYNSDLNTINTRVHEI 227

Qy 235 ANYVMYLYKLNTHVALVGMETWTKWKIKITPNASWTLENKSKRGSVLSRRKHHWIAQ 294
 Db 228 VNFINGFYRLNIHVSLTDLSEWQNEQINIQSASDTLNAFAEWRETDLLARKSHDNAQ 287

Qy 295 LITATELAGTTVGLAMKSTWMSR-YSVGVVQVHSMNLLRVAGTMAHEMGKNGKWHHWYS 353
 Db 288 LLTAIELDSETLGLAPLGTWCDDPKLS-GIVQDHPINLLMGVTVAHELGEWGMHHDGKD 347

Qy 354 WKPESTIWMKALSWYIPTWMSWSRSLSYKWKWENKLSNWLWNAPLPTWIIISTPIWGNQ 413
 Db 348 CLRGAELCIMPGLTKGRSYSEFSADSMHYERFLKQYKQPCILNKPLRIDPVPSTVPSGNE 407

Qy 414 LVEAGE 419
 Db 408 LLEAGE 413

RESULT 14

JC4880
 N:Altrinate names: lebetase Le3
 C:Species: Vipera lebetina
 C;Date: 10-Sep-1996 #sequence_revision 18-Oct-1996 #text_change 20-Jun-2000
 C;Accession: JC4880; PC4181
 R;Sigur, E.; Aaspollu, A.; Tu, A.T.; Sigur, J.
 Biochim. Biophys. Res. Commun. 224, 229-236, 1996
 A;Title: cDNA cloning and deduced amino acid sequence of fibrinolytic enzyme (lebetase)
 A;Reference number: JC4880; MUID:96280739; PMID:8694817
 A;Accession: JC4880
 A;Molecule type: mRNA
 A;Residues: 1-478 <SI11>
 A;Cross-references: EMBL:X97894; NID:gl502367; PIDN:CRA66471.1; PID:gl502368
 A;Accession: PC4181
 A;Molecule type: protein
 A;Residues: 214-246 <SI12>
 A;Experimental source: venom
 C;Superfamily: trigramin precursor; disintegrin homology
 C;Keywords: anticoagulant; glycoprotein; hydrolase; metalloproteinase; venom; zinc
 F;1-18/Domain: signal sequence #status predicted <SIG>
 F;19-191/Domain: activation peptide #status predicted <ACP>
 F;92-397/Product: fibrinolytic metalloproteinase #status predicted <MAT>
 F;404-474/Domain: disintegrin homology <DIS>
 F;416-475/Domain: disintegrin #status predicted <DNT>
 F;456-458/Region: cell attachment (R-G-D) motif
 F;170,337,341,347/Binding site: zinc, catalytic (Cys, His, His) (inhibited) #status
 F;312-392,352-376,354-359/Disulfide bonds: #status predicted
 F;337,341,347/Binding site: zinc, catalytic (His) (active) #status predicted
 F;338/Active site: Glu #status predicted

Query Match 11.4%; Score 536; DB 2; Length 478;

Best Local Similarity 35.0%; Pred. No. 1.6e-31;

Matches 148; Conservative 72; Mismatches 187; Indels 16; Gaps 8;

Qy 1 MLOGLLPVSLLSV-----AVSAIKELPGVKYKVVYPIRLHPLHKKRAKEPEQOEOWETE 56

Db 1 MIOVLL-VTICLAVFPYQGSIILESQVNDYEVVNPQAVTGLPKGAVKQPE--KKYEDT 57

Qy 57 LKYMRTINGIAVLVLYLKKNNKLLAPGYTETYYNSTGKEITTSPOIMWWWYQGHILNEKV 116
 Db 58 MOYEFVNGEPVVLHLEKKNELFSKDYSETHYSPDGREITINPAVEDHCYHGRIONDAD 117

Qy 117 SWASISTWRGLRGYSQGWQRYWIEPLSPIHRWQOEHALWKYNPWEKNYKYSTWGMWGLW 176
 Db 118 STASISACGLKGFTLRETLYLEPLKLPD--SEAHAVVYKYEKEDAP--KMGCVTQ 174

Qy 177 AHMLQQNIAPATKLVKLKWRKQOE--HEKYIEYLVLMNGENKRYNENOWEIRKRVWEM 234
 Db 175 TNNASDE---PIKKASQLNLTPEQOEFEPRIYELIVADHAMVTKYNGDLAAITTVHQL 231

Qy 235 ANYVMYLYKLNTHVALVGMETWTKWKIKITPNASWTLENKSKRGSVLSRRKHHWIAQ 294
 Db 232 VNNINGFYRLNVHITLSAVEVWINGDLINVPAASTVTLNLFGEWRERDLLNRKHDAQ 291

Qy 295 LITATELAGTTVGLAMKSTWMSR-YSVGVVQVHSMNLLRVAGTMAHEMGKNGKWHHWYS 353
 Db 292 LLTGIDLDNITIGLADDDSMCDPRYSYGVQDHSAILRLVAVTVAHELGHNLGNHHDGQ 351

Qy 354 WKPESTIWMKALSWYIPTWMSWSRSLSYKWKWENKLSNWLWNAPLPTWIIISTPIWGNQ 413
 Db 352 CNGAGCVMVSVLIEORSYQFSDCSKNKYOTYLTNRNPQCILNQLPLTDTTVPVSGNE 411

Qy 414 LVE 416
 Db 412 LIQ 414

RESULT 15

S43125
 trimucin precursor - Chinese habu
 N:Contains: hemorrhagic proteinase (EC 3.4.24.-); platelet aggregation inhibitor (dis:
 C;Species: Trimeresurus mucrosquamatus (Chinese habu)
 C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 22-Jun-1999
 C;Accession: S47570; S43125
 R;Tsai, I.H.; Wang, Y.M.; Lee, Y.H.
 Biochim. Biophys. Acta 1200, 337-340, 1994
 A;Title: Characterization of a cDNA encoding the precursor of platelet aggregation inh
 A;Reference number: S47570; MUID:94347824; PMID:8068721
 A;Accession: S47570
 A;Molecule type: mRNA
 A;Residues: 1-481 <TS2>
 A;Cross-references: EMBL:X77089; NID:9467703; PIDN:CAA54364.1; PID:9467704
 C;Superfamily: trigramin precursor; disintegrin homology
 C;Keywords: anticoagulant; hydrolase; metalloproteinase; venom; zinc
 F;1-18/Domain: signal sequence #status predicted <SIG>
 F;13-187/Domain: propeptide #status predicted <PRO>
 F;188-392/Product: hemorrhagic proteinase #status predicted <HEM>
 F;399-477/Domain: disintegrin homology <DIS>
 F;411-478/Product: disintegrin #status predicted <DNT>
 F;459-461/Region: cell attachment (R-G-D) motif
 F;333,337,343/Binding site: zinc, catalytic (His) #status predicted
 F;334/Active site: Glu #status predicted

Query Match 11.4%; Score 535; DB 2; Length 481;

Best Local Similarity 33.7%; Pred. No. 1.9e-31;

Matches 163; Conservative 88; Mismatches 193; Indels 40; Gaps 14;

Qy 1 MLOGLLPVSLLSV-----AVSAIKELPGVKYKVVYPIRLHPLHKKRAKEPEQOEOWETE 56

Db 1 MIOVLL-VTICLAVFPYQGSIILESQVNDYEVVYPAKVATLPGAV-----QPKYEDA 54

Qy 57 LKYMRTINGIAVLVLYLKKNNKLLAPGYTETYYNSTGKEITTSPOIMWWWYQGHILNEKV 116
 Db 55 MOYEFVNGEPVVLHLEKKNELFSKDYSETHYSPDGREITINPAVEDHCYHGRIONDAD 114

Qy 117 SWASISTWRGLRGYSQGWQRYWIEPLSPIHRWQOEHALWKYNPWEKNYKYSTWGMWGLW 176

Db 115 STASISACGLKGFKYKLGQETYPLEPLESD--SEAHAVFKEYENVEKDEAP--KMGCVT- 170

Qy 177 AHMLQQNIAPATKLVKLKWRKQOEHEKYIEYLVLMNGENKRYNENOWEIRKRVWEMAN 236

```

171 QWESDESIIKASQLY-LTPECCQPPQRYIKLAIUVDHGMYTKYSSNPKIKRVRHQMVS 229
237 YVNLVKKLINTHVALVGMIEIHTWKWKIKITPNASWTLENWSKRGSVLSRRKRHWIAQILI 296
230 NINEMCPFLNIAITLALIDVWSEKDFIVQADAPTAGLFGWRVRVLLKKKNHDHAQILL 289
297 TATELAGTTCVLAWKSTMW-SPYSVGVVQWHSNLLRVAGTVAHBMHGNWGMWHSWK 355
290 TTNFANTICISAYVGRMCDSEKYSVAVVVDHSSKVFXVAVTTHELGHNLGHEHDKD-K 348
356 WPSTIIVWYWKALSUYIPTWSSWRLS-----YKMWKWKLSN---WLNAPLPTWII 405
349 CKDTCIMSAVIS-----DKQSLSDSCSKDYQYTF---LTNDNPCCILNAPLRTDTV 398
406 STPIWKGQIEMGEHWWWWGTSEEWNTIWWAKTKIKATWQWALGEWKEWQWKAGWVW 465
399 STPYSGNEFEELAGEECDGSEF---NPCCDAATCKLRPGACQREGLCCDQCCKFKKRTIC 455
466 RPAK 469
456 RRAR 459

```

Search completed: June 9, 2004, 18:02:52
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2004, 17:51:20 ; Search time 17 Seconds
(without alignments)
2373.786 Million cell updates/sec

Title: US-10-726-148A-15

Perfect score: 4712

Sequence: 1 MLQGLLPVSLLSVAISAIAK.....PFTVKNPMSPTKWSNPKA 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3120	66.2	775	1	A028_HUMAN
2	2775.5	58.9	776	1	A028_MACFA
3	2071.5	44.0	793	1	A028_MOUSE
4	773.5	16.4	776	1	A007_MACFA
5	764.5	16.2	788	1	A007_MOUSE
6	755.5	16.0	789	1	A007_RAT
7	750	15.9	754	1	A007_HUMAN
8	685.5	14.5	826	1	A008_MOUSE
9	671	14.2	571	1	D1S3_BOTJA
10	642	13.6	824	1	ECAR_ECHCA
11	633	13.4	616	1	ECAR_ECHCA
12	574	12.2	483	1	D1S3_TRIFL
13	543	11.5	414	1	H1D1_CROAT
14	539.5	11.4	956	1	A019_HUMAN
15	534.5	11.3	478	1	H1E1_CROAT
16	529.5	11.2	413	1	ACLA_AGRAC
17	516	11.0	480	1	D1S3_TRIGA
18	501	10.6	920	1	A019_MOUSE
19	484	10.3	478	1	D1S3_AGRKH
20	463.5	9.8	797	1	A033_MOUSE
21	463.5	9.8	909	1	A012_HUMAN
22	462.5	9.8	813	1	A033_HUMAN
23	457.5	9.7	903	1	A012_MOUSE
24	436.5	9.3	416	1	H1B1_TRIFL
25	432	9.2	401	1	KAOU_NAJKA
26	421	8.9	819	1	A009_HUMAN
27	401.5	8.5	814	1	A015_HUMAN
28	373	7.9	864	1	A015_MOUSE
29	363.5	7.7	816	1	A015_RAT
30	337.5	6.7	697	1	A026_MOUSE
31	309.5	6.6	726	1	A020_HUMAN
32	296.5	6.3	201	1	H1B1_TRIFL
33	284.5	6.0	200	1	H1B1_LACMU

34	281.5	6.0	202	1	H1B1_CROAT
35	276.5	5.9	735	1	A002_CAVPO
36	271.5	5.8	729	1	A002_MOUSE
37	271	5.8	751	1	A002_RABIT
38	268.5	5.7	202	1	H1B1_TRIFL
39	265.5	5.6	203	1	F1B1_AGRAC
40	265.5	5.6	760	1	A025_MOUSE
41	265	5.6	954	1	XVNA_RUMEL
42	262.5	5.6	203	1	A00A_CROAD
43	257	5.5	789	1	A001_RAT
44	257	5.5	935	1	A022_XENLA
45	254.5	5.4	820	1	A029_HUMAN

ALIGNMENTS

RESULT 1
AD28_HUMAN
ID AD28_HUMAN STANDARD; PRT; 775 AA.
AC Q9UKQ2: Q9Y339; Q9Y350;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE ADAM 28 precursor (EC 3.4.24.-) [A disintegrin and metalloproteinase domain 28] (Metalloproteinase-like, disintegrin-like, and cysteine-rich protein-L) (MDC-L) (eMDC II) (ADAM23).
GN ADAM28 OR MDCL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Lymph node;
RX MEDLINE=99436132; PubMed=10506182;
RA Roberts C.M., Tani P.H., Bridges L.C., Kaszik Z., Bowditch R.D.;
RT "MDC-L, a novel metalloproteinase disintegrin cysteine-rich protein family member expressed by human lymphocytes.";
RL J. Biol. Chem. 274:29251-29259(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Epithelium;
RX MEDLINE=20054911; PubMed=10587367;
RA Jary J.A., Perry A.C., Hall L.;
RT "Identification, sequence analysis and expression of transcripts encoding a putative metalloproteinase, eMDC II, in human and macaque epithelium.";
RL Mol. Hum. Reprod. 5:1127-1134(1999).
CC -!- FUNCTION: May play a role in the adhesive and proteolytic events that occur during lymphocyte emigration or may function in ectodermal shedding of lymphocyte surface target proteins, such as FASL and CD40L. May be involved in sperm maturation.
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (isoform 1). Secreted (isoform 2).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=MDC-LM;
CC IsoId=Q9UKQ2-1; Sequence=Displayed;
CC Name=2; Synonyms=MDC-LI;
CC IsoId=Q9UKQ2-2; Sequence=VSP_005487;
CC -!- TISSUE SPECIFICITY: Expressed predominantly in secondary lymphoid tissues, such as lymph node, spleen, small intestine, stomach, colon, appendix and trachea. The lymphocyte population is responsible for expression of this protein in these tissues.
CC Isoform 2 is expressed preferentially in spleen.
CC -!- PTM: Pro-domain removal and maturation may be, at least in part, autocatalytic (By similarity).
CC -!- SIMILARITY: Belongs to peptidase family M12B.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC -!- SIMILARITY: Contains 1 disintegrin domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL	AF137334	AD25099.1	-	
DR	EMBL	AF137335	AD25100.1	-
DR	EMBL	AU242015	CAB42085.1	-
DR	HSSP	P18619	1FVL	
DR	MEROPS	M12.224	-	
DR	Genes	HGNC:206	ADAM28	
DR	MIM	606186	-	
DR	GO	GO:0008237	F-metalloproteinase activity	TAS
DR	GO	GO:0007283	P-spermatogenesis	TAS
DR	InterPro	IPR006586	ADAM cysteine	
DR	InterPro	IPR001762	Disintegrin	
DR	InterPro	IPR006209	EGF-like	
DR	InterPro	IPR006210	IEGF	
DR	InterPro	IPR001818	Pept_M10A_M12B	
DR	InterPro	IPR006025	Pept_M_Zn_BS	
DR	InterPro	IPR001590	Peptidase_M12B	
DR	InterPro	IPR002870	Peptidase_M12B_N	
DR	Pfam	PF00200	disintegrin	1
DR	Pfam	PF01562	Pept_M12B_propep	1
DR	Pfam	PF01421	Retrolysin	1
DR	PRINTS	PR00289	DISINTEGRIN	
DR	ProDom	PD00664	Disintegrin	1
DR	SMART	SM00608	ACR	1
DR	SMART	SM00050	DISIN	1
DR	SMART	SM00181	EGF	1
DR	PROSITE	PS00215	ADAM_MPRO	1
DR	PROSITE	PS00546	CYS-STEIN SWITCH	FALSE_NEG
DR	PROSITE	PS00427	DISINTEGRIN_1	FALSE_NEG
DR	PROSITE	PS00214	DISINTEGRIN_2	1
DR	PROSITE	PS00022	EGF_1	FALSE_NEG
DR	PROSITE	PS01186	EGF_2	1
DR	PROSITE	PS00026	EGF_3	1
DR	PROSITE	PS00142	ZINC_PROTEASE	1
DR	KW	Hydrolase	Metalloprotease	Zinc
DR	KW	Transmembrane	EGF-like domain	Alternative splicing
FT	SIGNAL	1	18	POTENTIAL
FT	PROPEP	19	198	BY SIMILARITY
FT	CHAIN	199	775	ADAM 28
FT	DOMAIN	199	665	EXTRACELLULAR (POTENTIAL)
FT	TRANSMEM	666	686	POTENTIAL
FT	DOMAIN	687	775	CYTOPLASMIC (POTENTIAL)
FT	DOMAIN	199	399	METALLOPROTEASE
FT	DOMAIN	407	493	DISINTEGRIN-LIKE
FT	DOMAIN	494	628	CYS-RICH
FT	DOMAIN	629	657	EGF-LIKE
FT	SITE	169	169	CYSTINE SWITCH (POTENTIAL)
FT	METAL	339	339	ZINC (CATALYTIC) (BY SIMILARITY)
FT	ACT SITE	340	340	BY SIMILARITY
FT	METAL	343	343	ZINC (CATALYTIC) (BY SIMILARITY)
FT	METAL	349	349	ZINC (CATALYTIC) (BY SIMILARITY)
FT	DISULFID	315	394	BY SIMILARITY
FT	DISULFID	354	378	BY SIMILARITY
FT	DISULFID	465	478	POTENTIAL
FT	DISULFID	629	639	BY SIMILARITY
FT	DISULFID	633	645	BY SIMILARITY
FT	DISULFID	647	656	BY SIMILARITY
FT	CARBOHYD	268	268	N-LINKED (GLCNAC...) (POTENTIAL)
FT	CARBOHYD	275	275	N-LINKED (GLCNAC...) (POTENTIAL)
FT	CARBOHYD	557	557	N-LINKED (GLCNAC...) (POTENTIAL)
FT	CARBOHYD	602	602	N-LINKED (GLCNAC...) (POTENTIAL)
FT	CARBOHYD	628	628	N-LINKED (GLCNAC...) (POTENTIAL)
FT	VARSPLIC	524	540	TEVADSCVNRNEGSK -> RRTNPFPCAKENHF (in isoform 2)
FT	VARSPLIC	541	775	Missing (in isoform 2)

FT				/FTId=VSP_005487.	
FT	CONFLICT	513	513	R -> Q (IN REF. 2).	
FT	CONFLICT	774	774	E -> K (IN REF. 2).	
SQ	SEQUENCE	775 AA;	87208 MW;	4A56359014DF97AL CRC64;	
Query Match 66.2%; Score 3120; DB 1; Length 775;					
Best local similarity 84.1%; Pred.No. 8.4e-221;					
Matches 652; Conservative 31; Mismatches 92; Indels 0; Gaps 0;					
Qy	1	MLOGGLPVSLLLSVAYS	AIKELPGVKYKYEVVYPIRLHPLHKRAKEPEQEOWETELKYK	60	
Dd	1	MLOGGLPVSLLLSVAIS	AIKELPGVKYKYEVVYPIRLHPLHKRAKEPEQEOPETELKYK	60	
Qy	61	MTINGIAVLVLYKKKNLAP	GVTYTETYNSTGKEITTSPOIMMHWYYQGHIANEKVSWAS	120	
Dd	61	MTINGIAVLVLYKKKNLAP	GVTYTETYNSTGKEITTSPOIMDCYQGHILNEKVSAS	120	
Qy	121	ISTWRGLRGYSQGWRYMEI	PLEPSPIHRWQOEHALWKYPWEKYNYSTWCVMGVLAHL	180	
Dd	121	ISTCRGLRGYSQDQRYFIE	PLEPSPIHRDQOEHALFKYNPDEKXYSTCGMDGVLAHDL	180	
Qy	181	QQNIALPATKLVLKRWKVOE	HEKYTEYLVLNMGEMKRYENQWEIRKEVPMANYVM	240	
Dd	181	QQNIALPATKLVLKDKRVOE	HEKYTEYLVLNNGEKRYENQDEIRKVPFMANVYM	240	
Qy	241	LYKKLNTHVALVGMEIWTW	KWKLIKITPNASWTILENNASKWSVLRRKHIIAQLITATE	300	
Dd	241	LYKKLNTHVALVGMEIWT	DKDIKITPNASFITLENFSKWEGSVLSRRKHHDIAQLITATE	300	
Qy	301	LACTTVGLAMNSTWNSPY	SGVGTQWISNLNLRVAGTMAHECHGNWGMHWYSHKWPKSTI	360	
Dd	301	LACTTVGLAFMSTMCSFY	SGVGYQDSNDLLRVAGTMAHECHGNFGHFHDDYSCCKCFSTI	360	
Qy	361	VWVKKALSIVPTPWSSWS	RLSKYKWKWKSNNLWNAPELCMIISTPIMGNOLVEMGEW	420	
Dd	361	CVMDKALSIFYIPTDFS	CSRSLSDXKFEDKLSNCLFNAPLPDTLIISTPICGNQJVEMGD	420	
Qy	421	WWWGTSEENTNIWWAAKT	NKIKAATWQALGBEWKWKQWKAGWVRPAKMWLPEMNWG	480	
Dd	421	CDGGTSEECTNICDAKT	CKIKATFOCALGECCEKQFKAGWVCRPFAKDECCLPEMCNG	480	
Qy	481	KSGNWPNRWONGVPHWH	HCKGHNLTGMTLEOQNTWLGPCTEVANKSWYRNESGSK	540	
Dd	481	KSGNCPPDRFQVNGFP	CHHGKGHCMLGTCPTRSEQCTLWGPCTEVADKSCYRNESGSK	540	
Qy	541	YGWVRVAVNTLIPKANWT	MWGLMWOGGSWNLPWKGRIVTWLTWTKTWPEWTSQEEGVY	600	
Dd	541	YGTCRRVDTLIPCKAND	TWCGKLFQCGASDNLPWKGRIVTFCTFDPTSQEEIGMV	600	
Qy	601	ANGTKGWNKVMVINAEM	WIERYAKYSTNNASSKWKGHAVMMHELQWMBEGSWIPPWWMS	660	
Dd	601	ANGTKGDGNKVCINAE	CVDIEKAYKSTNCSSCKCKGHAVCDHELQCCEBGIWPFDDCDS	660	
Qy	661	VVWHSAIVGVLPMAVIW	VVVAMVIRHOSREKQKKQBPLSTTGTRPHKQRKPOMVK	720	
Dd	661	WFHFSIVGVLPMAVIF	VVWAMVIRHOSREKQKKQORPLSTTGTRPHKQRKPOMVK	720	
Qy	721	AVQPQEMSQMKBHVWLP	PVBGNEPPASMKHTNALPPTYVKKNPMSTPKWMSNKA	775	
Dd	721	AVQPQEMSQMKBHYDLP	PVBGNEPPASFKHTNALPPTYVKKNPMSTPKDSNEFA	775	
RESULT 2					
AD28_MACEFA					
ID	_AD28_MACEFA	STANDARD;	PRT;	776 AA.	
AC	Q9XSU6;				
DT	16-OCT-2001 (Rel. 40,	Created)			
DT	16-OCT-2001 (Rel. 40,	Last sequence update)			
DT	28-FEB-2003 (Rel. 41,	Last annotation update)			
DE	ADAM 28 precursor (EC	3.4.24.-) (A disintegrin and metalloproteinase			
DE	domain 28) (emDC II).				
GN	Macaca fascicularis (Crab eating macaque) (Synomolgus monkey).				
GN	ADAM28.				

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca.
 OK NCBI_TaxID=9541;
 [1]_SEQUENCE FROM N.A.
 RP TISSUE=Epididymis;
 RC MEDLINE=20054911; PubMed=10587367;
 RX JURY J.A., PERRY A.C., HALL L.;
 RA "Identification, sequence analysis and expression of transcripts
 RT encoding a putative metalloproteinase, eMDC II, in human and macaque
 RL epididymis";
 RL Jol. Hum. Reprod. 5:1127-1134(1999).
 CC -!- FUNCTION: May play a role in the adhesive and proteolytic events
 CC that occur during lymphocyte emigration or may function in
 CC ectodermal shedding of lymphocyte surface target proteins, such as
 CC PASL and CD40L. May be involved in sperm maturation.
 CC -!- COPACTOR: Binds 1 zinc ion per subunit (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: Expressed at high levels in epididymis and at
 CC lower levels in lung.
 CC -!- PTM: Pro-domain removal and maturation may be, at least in part,
 CC autocatalytic (By similarity).
 CC -!- SIMILARITY: Belongs to peptidase family M12B.
 CC -!- SIMILARITY: Contains 1 EGF-like domain.
 CC -!- SIMILARITY: Contains 1 disintegrin domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AJ242014; CAB42090.1; -
 DR HSSP: P18619; IFVL.
 DR MEROPS: M12.224; -
 DR InterPro: IPR006586; ADAM cysteine.
 DR InterPro: IPR001762; Disintegrin.
 DR InterPro: IPR006209; EGF like.
 DR InterPro: IPR006210; IEGF.
 DR InterPro: IPR001818; Pept_M10A_M12B.
 DR InterPro: IPR006025; Pept_M_Zn_BS.
 DR InterPro: IPR001590; Peptidase_M12B.
 DR InterPro: IPR002870; Peptidase_M12B_N.
 DR Pfam: PF00200; disintegrin; 1.
 DR Pfam: PF01562; Pept_M12B_propep; 1.
 DR Pfam: PF01421; Redolyein; 1.
 DR PRINTS: PR00289; Disintegrin.
 DR ProDom: PD000664; Disintegrin; 1.
 DR SMART: SM00608; ACR; 1.
 DR SMART: SM00050; DISIN; 1.
 DR SMART: SM00181; EGF; 1.
 DR PROSITE: PS00215; ADAM_MEPRO; 1.
 DR PROSITE: PS00546; CYSTEINE_SWITCH; FALSE_NEG.
 DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
 DR PROSITE: PS00214; DISINTEGRIN_2; 1.
 DR PROSITE: PS00022; EGF_1; FALSE_NEG.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS00026; EGF_3; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 KW Transmembrane; EGF-like domain.
 FT SIGNAL 1 19 POTENTIAL.
 FT PROPEP 20 193 BY SIMILARITY.
 FT CHAIN 194 776 ADAM 28.
 FT DOMAIN 194 666 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 667 687 POTENTIAL.
 FT DOMAIN 688 776 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 194 400 METALLOPROTEASE.
 FT DOMAIN 408 494 DISINTEGRIN-LIKE.
 FT DOMAIN 495 629 CYS-RICH.

FT	DOMAIN	625	658	EGF-LIKE.
FT	SITE	170	170	CYSTEINE_SWITCH (POTENTIAL).
FT	METAL	340	340	ZINC (CATALYTIC) (BY SIMILARITY).
FT	ACT_SITE	341	341	BY SIMILARITY.
FT	METAL	344	344	ZINC (CATALYTIC) (BY SIMILARITY).
FT	METAL	350	350	ZINC (CATALYTIC) (BY SIMILARITY).
FT	DISULFID	315	395	BY SIMILARITY.
FT	DISULFID	355	379	POTENTIAL.
FT	DISULFID	466	479	BY SIMILARITY.
FT	DISULFID	630	640	BY SIMILARITY.
FT	DISULFID	634	646	BY SIMILARITY.
FT	CARBOHYD	268	268	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	275	275	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	352	352	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	558	558	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	603	603	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	629	629	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	776 AA;	87213 MW;	08AAPE834B37F19F CRC64;

Query Match 58.9%; Score 2775.5; DB 1; Length 776;
 Best Local Similarity 76.1%; Pred. No. 1.3e-195;
 Matches 596; Conservative 49; Mismatches 123; Indels 15; Gaps 6;

QY	1	MLQGLPVSLLS-VAVSAIKELPGVKYEVVYPIRLPLHFKREAKEPEQOQWETELKY	59
DB	1	MLQALLTVSLLSVPVSAIKELPGVKYEVVYPIRLPLHFKREAKEPEQOQWETELKY	60
QY	60	KMTINGKIAVLKKNKLLAPGTYTYNSTGKEITTSQIMMWWYYQHILNEKWSNA	119
DB	61	KMTWNGKIAVLKKNKLLAPGTYTYNSTGKEITTSQIMDDCYQGHIIINEKSDA	120
QY	120	SISTWGLRGYQGWRYIELSPHRCQEHALWKYAPWKNYVSTWCMGWGLWAHW	179
DB	121	SISTCGLRGYQGNQRYFIELSPHRCQEHALWKYAPWKNYVSTWCMGWGLWAHW	180
QY	180	LQONIALPATKLVKWKVKQHEKYLEYVLVWNGEWKRYNEQWEIRKRVEMANYN	239
DB	181	L-QNIARPATLVKLVNDGKVKQHEKYLEYVLVWNGEFKKYNEQDEIRKRVEMANYN	239
QY	240	MLYKCLMTHVALVGMETWTKWKIKITPNASWTLENKSKRGSVLSRKRHWIAQLITAT	299
DB	240	MLYKCLMTHVALVGMETWTDKINITENASFLENFSKWRGSLPRKRKHIAQLITAT	299
QY	300	ELAGTTVGLAWMTWMSPY-SVGVVVOMHSNKLRLVAGTMAHEMGNHGMHWTSWKWPS	358
DB	300	EFAGMTVGLAFPMSTMCSFYHSVGVDHSONLLRVAGTMAHEMGNHGMHFNHNSCKCPS	359
QY	359	TIWVMKALSWYIPTWSSWSRLSYKWKWVKLNLWNAELPTWIIISTPIWGNQLVBMG	418
DB	360	TICVMDKALSEYIPTDFSSCSRVSVDKFFEDKLSNCLFNAFLPTDIIISTPICGNQWBMG	419
QY	419	EMWVGTSSEWTNIWMAKTKIKATWCMALGEWKEKWKAGWVPAKWKWLPKMW	478
DB	420	EDDCGTSSECTNICDAKTKIKAGEQTSSECECKQFKKAGWVCPAKWDECDLPMC	479
QY	479	NKSGSNPWVWQVNGVPHGKHLWGTWPTLQEOMTLWPGTETVAMKSWYNRNKG	538
DB	480	DGKSGNCPDDFRFRANGPPCHGKGYCLMGACPTLQEOCTELWPGTKVADQSCYNRNGG	539
QY	539	SKYGYRVRVWMTLIPWKNMTWGLKMWQGSWMLPKWGRIVTWLTWTKWPEWTSQBIG	598
DB	540	SKYGYCRVDDTLIPCKTNDTMCGLFCQGSNDLPMWGRIVTWLTCTFPDPTSEBIG	599
QY	599	MYVANGTKWGMKNWIAHWWIEKAYKSTNNSKWKHAWVHLOWEGERWIPWVWV	658
DB	600	MYVANGTKGHNKVCINAEVDIEKAYKSTNNSKWKHAWVHLOWEGERWIPWVWV	659
QY	659	SSVWVNSIVVGLWPMVAIVVWVAVIRHOSSEKOKKORP-STTCTTRPHKQKPKQM	718
DB	660	SSVVFYSIVVAVLFPVAVISLVVAIVIRQOSSREKQKQORP-STTCTTRPHKQKPKQM	719
QY	719	VKAVQPMQMSQMKPHVYVWLPVEGNEPPASW-HKWTNALPPTVWKNPMSTPKW-----SN	772

```
Db 720 VXAQPMQSMKLVHVDLVEGNEPASPFLSKPDSFP-----PIPAKSSSFLDSN 773
QY 773 PKA 775
DB 774 PKA 776

RESULT 3
AD28 MOUSE
ID AD28 MUISB STANDARD; PRT: 793 AA.
AC Q9JUN6; Q9K5D2; Q8K5D3;
DT 16-OCT-2001 (rel. 40, Created)
DT 28-FEB-2003 (rel. 41, last sequence update)
DT 10-OCT-2003 (rel. 42, last annotation update)
DE ADAM 28 precursor (BC 3.4.24.-) (A disintegrin and metalloproteinase
DE domain 28) (thymic epithelial cell-ADAM) (TECADAM).
GN ADAM28.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxid=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 3), CHARACTERIZATION, AND MUTAGENESIS
RP OF GLU-343.
RC TISSUE=lung;
RX MEDLINE=20256759; PubMed=10794709;
RA Howard L., Maciewicz R.A., Biobel C.P.;
RT "Cloning and characterization of ADAM28: evidence for autocatalytic
RT pro-domain removal and for cell surface localization of mature
RT ADAM28."
RL Biochem. J. 348:21-27(2000).
RK [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3), AND TISSUE SPECIFICITY.
RC TISSUE=Thymus;
RX MEDLINE=21856493; PubMed=11867223;
RA Haidl I.D., Huber G., Eichmann K.;
RT "An ADAM family member with expression in thymic epithelial cells and
RT related tissues."
RL Gene 283:163-170(2002).
CC -!- FUNCTION: May play a role in organogenesis and organ-specific
CC functions such as thymic T cell development.
CC -!- CAPTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=Q9JUN6-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9JUN6-2; Sequence=VSP_005488;
CC Name=3;
CC IsoId=Q9JUN6-3; Sequence=VSP_005489, VSP_005490;
CC -!- TISSUE SPECIFICITY: Strong expression in thymic epithelial cells
CC and developmentally related tissues including the trachea,
CC thyroid, lung and stomach, but not in lymphocytes. Expressed at
CC high levels also in epididymis. In contrast with human is not
CC expressed in immature or mature lymphocyte populations of
CC thymocytes, lymph node, spleen, and bone marrow.
CC -!- DEVELOPMENTAL STAGE: The expression patterns in adult and day 15.5
CC embryos are similar.
CC -!- PTM: Pro-domain removal and maturation may be, at least in part,
CC autocatalytic.
CC -!- SIMILARITY: Belongs to peptidase family M12B.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC -!- SIMILARITY: Contains 1 disintegrin domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/).
```

```
or send an email to license@isb-sib.ch.
CC
DR EMBL; AF153350; RA071993.1; -
DR EMBL; AF163290; AAM21935.1; -
DR EMBL; AF163291; AAM21936.1; -
DR EMBL; AF163292; AAM21937.1; -
DR HSSP; P18619; 1FVL.
DR MEROPS; M12.020; -
DR MGD; MGI:105988; Adam28.
DR InterPro; IPR006586; ADAM cysteine.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR001818; Pept_M12B.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001590; Peptidase M12B.
DR InterPro; IPR002870; Peptidase M12B_N.
DR Pfam; PF00200; Disintegrin; 1.
DR Pfam; PF01562; Pept_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR PRINTS; PR00289; DISINTEGRIN.
DR ProDom; PD003664; Disintegrin; 1.
DR SMART; SM00608; ACR; 1.
DR SPART; SM00050; DISIN; 1.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00546; CYSTEINE_SWITCH; FALSE NEG.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE NEG.
DR PROSITE; PS00214; DISINTEGRIN_2; 1.
DR PROSITE; PS00022; EGF_1; FALSE NEG.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS50026; EGF_3; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KM Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KM Transmembrane; EGF-like domain; Alternative splicing.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 21 195 BY SIMILARITY.
FT CHAIN 196 793 ADAM 28.
FT DOMAIN 196 668 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 669 689 POTENTIAL.
FT DOMAIN 690 793 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 198 402 METALLOPROTEASE.
FT DOMAIN 410 496 DISINTEGRIN-LIKE.
FT DOMAIN 497 631 CIS-RICH.
FT DOMAIN 628 660 POLY-PRO.
FT SITE 171 171 EGF-LIKE.
FT METAL 342 342 CYSTEINE SWITCH (POTENTIAL).
FT ACT_SITE 343 343 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 346 346 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 352 352 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 317 397 BY SIMILARITY.
FT DISULFID 357 381 BY SIMILARITY.
FT DISULFID 468 481 POTENTIAL.
FT DISULFID 632 642 BY SIMILARITY.
FT DISULFID 636 648 BY SIMILARITY.
FT DISULFID 650 659 BY SIMILARITY.
FT CARBOHYD 91 91 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 277 277 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 531 531 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 551 551 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 605 605 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPPLIC 769 793 Missing (in isoform 2).
FT VARSPPLIC 769 771 /FTId=VSP_005488.
FT VARSPPLIC 775 793 TGR -> DPN (in isoform 3).
FT VARSPPLIC 775 793 /FTId=VSP_005489.
FT MUTAGEN 343 343 Missing (in isoform 3).
FT SEQUENCE 793 AA; 88670 MW; 7715E71456D403B CRC64;
SQ
Query Match 44.0%; Score 2071.5; DB 1; Length 793;
Best Local Similarity 60.3%; Pred. No. 4.2e-144;
```


Db 169 -----SCVGLNFTKKSALIDVE--NTEENAEADHKEKTELFVVADEYVYRNNKPN 220
QY 226 EIRKRYEMANYNMLYKKNLTHVALGMEITWKKIKITPNASWTLENKWSKRGSVLS 285
Db 221 KLRLNGIWMGVFNMYIKTLNTHVALGMEITWKKIKITPNASWTLENKWSKRGSVLS 285
QY 286 RYKRWIAQLITATELAGTITVGLAMSTMWSPYSVGVVWHSNLLR----- 332
Db 281 KRDFDHVILLGSKWLYTSMOGIA-----YPGGICQ-----ILRSCSVKDLDPDVN 327
QY 333 -VAGTMAHEGNGWGWYSHKWNSTIWMKALSWIPT-WGSSWSRLSYKWKWEMK 390
Db 328 IIGNRWVHGLHSLGMOHDFPCTCLGKCVMDG---SIPALKFKCSGTQYQQLDQ 384
QY 391 LSNWLNAPLPTWILSTPTWGNQVEMGEMWMTSEETNIWAKTWIKATWQWALG 450
Db 385 XPACILNPFPEKENDYFPCGNKKVDEGEKDCGPPVQECTNCCDAHKCVLKPGETCVES 444
QY 451 EWKWKQWKKAGWVRPAKEMWLLPEMNGKSGNFWPWRQVNGWFWHGHGHLWLGTPW 510
Db 445 ECECCQMKKEGAVCLARKNECDISEVCTSYSPCEPKDEFQANGPCRNKGYCFMGLCP 504
QY 511 TLQEQWTELWGPCTEVAWKNRNEGGSKYWRVWTLIPWKNWMTMGLWQGG- 569
Db 505 TRNEOCSELFITGGAESHSLCYRMKNKGNRFYCKNGKNTFVPCBEKDLCKCKIYCSGR 564
QY 570 -----SNWLPWKGRIVTWLTKWPEWTSOIEIGWANGTWGNWKNWYNABWVI 620
Db 565 PSSRLGENKTYLNKVKQNT-IKRTMPLHNSRDMGLVNSGTGCGDMVCSGECIEM 623
QY 621 EKAKSTWSSKWKGVHAWHELOKQWEEGIPPPWMSVWVWHSIVGVWLPMAVIV 680
Db 624 EKAYNSTICSPCDENDVDDNEPECCQCEGSIITWGEALNTSVINIVL--VWVLEV 681
QY 681 -VWAVIRHOSRRKQKQKQPLSTGT-----RPHKQKRP 716
Db 682 GLVILLIRYQKIKKQVQSSPREIRGVEMKGYFPEHQTRSEP 725

RESULT 6
AD07 RAT
ID AD07 RAT
AC STANDARD; PRT: 789 AA.
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE ADAM 7 precursor (A disintegrin and metalloproteinase domain 7)
GN ADAM7 OR EAPI.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE OF 26-38.
RX MEDLINE=93038491; PubMed=1417724;
RA Perry A.C.F., Jones R., Barker P.J., Hall L.;
RT snake venom haemorrhagic peptides.";
RL Biochem. J. 286:671-675 (1992).
CC -!- FUNCTION: May play an important role in male reproduction
CC including sperm maturation and gonadotropin function. This is a non
CC catalytic metalloproteinase-like protein (by similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Expressed specifically in the caput region of
CC the epididymis.
CC -!- SIMILARITY: Belongs to peptidase family M12B.
CC -!- SIMILARITY: Contains 1 disintegrin domain.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch.

DR EMBL; X66140; CAA46930.1; -
DR F1R; S28259; S28259.
DR HSP; P18619; IFVL.
DR MEROPS; M12.956; -
DR InterPro; IPR006586; ADAM_cysteine.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR001590; Peptidase_M12B.
DR InterPro; IPR002870; Peptidase_M12B_N.
DR Pfam; PF00200; disintegrin; 1_M12B_N.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR PRINTS; PR00289; Disintegrin.
DR ProDom; PD000664; Disintegrin; 1.
DR SMART; SM00608; ACR; 1.
DR SMART; SM00050; DISIN; 1.
DR PROSITE; PS0215; ADAM_MEPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; 1.
DR PROSITE; PS0214; DISINTEGRIN_2; 1.
KW Signal; Glycoprotein; Transmembrane.
FT SIGNAL 1 25
FT PROPEP 26 176 POTENTIAL.
FT CHAIN 177 789 ADAM 7.
FT DOMAIN 177 668 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 669 689 POTENTIAL.
FT DOMAIN 690 789 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 177 400 METALLOPROTEASE-LIKE.
FT DOMAIN 401 487 DISINTEGRIN-LIKE.
FT DOMAIN 488 668 CVS-RICH.
FT DISULFID 310 388 BY SIMILARITY.
FT DISULFID 350 372 BY SIMILARITY.
FT DISULFID 459 472 POTENTIAL.
FT CARBOHYD 84 84 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 167 167 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 583 583 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 628 628 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 664 664 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 789 AA; 89362 MW; EF1E7F1C5EF0779 CRC64;
Query Match 16.0%; Score 755.5; DB 1; Length 789;
Best Local Similarity 32.2%; Pred. No. 7.6e-48;
Matches 236; Conservative 141; Mismatches 285; Indels 71; Gaps 21;
QY 5 LLPVSL-LLSVAVSAI--KELPGVKYEVVYVIRLHPLHKEAK---EPEQEQWETELK 58
Db 1 MFPTGIFLMSVLISQMGQGVGVGQELVHPKLSLQKEDLERHDSOTFEEYEBELL 60
QY 59 YKMTINGKIAVLKKNKLLAPGYTETYNSTGKEITTSPOIMMMWYQGHILNEKVS 118
Db 61 YEIKGKKTUHLKAREFLALNYSETYINIKREMTVRHPQLDHCFCQSSIIHFDSA 120
QY 119 ASISTWRGLRGVWSQGWQVIRPLSPIHRWQOE--HALWKYN---PWKQYVSTWQMG 173
Db 121 ASISTCGLRGVPRVNDQRYLIEPV---KISDEGHLVFKYVNVKAPYATNY--SCEGL-- 173
QY 174 VLWAHLQONIALPATKVLKWKVQOEH--EKYIEYVLVLMNGEKVYVNVQWETKRV 231
Db 174 ----NFTKSTLIDA-KII--BEHKVEDVHKEFIELFVVADEYVYRNNKPNKRI 226
QY 232 WEMANYVNLKLVKLVTHVALGMEITWKKIKITPNASWTLENKWSKRGSVLSRKRH 291
Db 227 WGVVFNMYIKALNIRVTLTGMEIWSAGDEIETVSNLESTLILHFSTWQETLKKKDF 286
QY 292 IAQLITATELAGTITVGLAMSTMWSPYSVGVVQ--WHSNLLR-----VAGTMAHEMG 342
Db 287 HVLLASGRWLYTSMOGIA-----YPGGICQTLRSCSVKDLDPDVNIIGNMAHQLG 338


```

QY 343 HNWGWHTWYHGWKNDSTIWMKALSWIPT-WSSWSRLSYKWKWKLNLWNAFLP 401
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 339 HSLGWRHDDPCTCPLGKCVMS--AGSIPAIKFKSCSQTQVQFLKWKQKACILNPLP 395
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 402 TWIISTPIWGHOLVEMGEMWMTSEETNIWRAKTWIKIATWQALGEWKEWQWKA 461
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 396 EEFNDYDFGKNKVDSEECGPGVQBCCTNCCDAHKCVLXPGFTCVGECECSCQMKKE 455
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 462 GMYRPAKWEWLPMPMNGKSGMPWWQVNGWPHHGGHGLMGTWPTLQEQWTELWG 521
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 456 GVICRPAKNECDISEVCTGYSPCKDESCANGPCKNGEGYCPMGLCPTRDDQCAELFS 515
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 522 PGTEVAMKSWNRNKGSKYGVNRVWTLIPKANTWMTMGKLMQGG-----SW 571
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 516 GGAESHSILVRNQRNRFYCKNDKNTFYPCEDKDKCGKIYCTGGRSAHLGEDKTY 575
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 572 NLPWKGRIVTWLTKTWPEWTSOISGVANGTKWGNKWYNABWVIEKAYKSTWSS 631
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 576 NLKNVKQNIS-ICKTMYLHNSDMLVNSGTHKCGSGVCSNGECIBEMKAYNSTICSS 634
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 632 KWKGHAVWHELQWQEGWIPPMWWSVVMHMSIVGVLPWMAVWV-VWAVIRHQ- 689
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 635 LCDENVDVNEPPCQCEBGIITWEGSMLNUTSVSINVVL-VNVLIGVGLVILLIRYQK 693
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 690 -----SSRE 693
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 694 CIOKQVQSSRE 706
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
AD07 HUMAN
ID AD07 HUMAN STANDARD; PRT; 754 AA.
AC C942U9; 075959;
DT 28-FEB-2003 (Rel. 41, Created;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ADAM 7 precursor (A disintegrin and metalloproteinase domain 7)
DE (Sperm maturation-related glycoprotein GP-83).
OS ADAM7 OR GP83
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1] _
RC TISSUE-Epithidymis;
RA Lin Y.C.; Lee Y.M.; Sun G.H.; Liu H.W.;
RT "Expression of human ADAM7 (GP-83).";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE OF 167-754 FROM N.A.
RP TISSUE-Epithidymis;
RC TISSUE-Epithidymis;
RA Lin H.W.; Lin Y.C.; Sun G.H.;
RT "A ADAM-like cDNA sequence identified in cDNA library of human
epithidymis.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May play an important role in male reproduction
including sperm maturation and gonadotrope function. This is a non
catalytic metalloproteinase-like protein (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: Belongs to peptidase family M12B.
CC -!- SIMILARITY: Contains 1 disintegrin domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement. See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch.
-----
EMBL; AF215824; AAC43987.1; -
EMBL; AF090327; AAC36742.1; -

```

```

DR HSSP; P18619; 1FVL.
DR Genew; HGNC:214; ADAM7.
DR MIM; 607310; -.
DR InterPro; IPR006586; ADAM cysteine.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR001590; Peptidase M12B.
DR InterPro; IPR002870; Peptidase M12B_N.
DR Pfam; PF00200; disintegrin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR PRINTS; PR00289; DISINTEGRIN.
DR ProDom; PD000664; Disintegrin; 1.
DR SMART; SM00608; ACR; 1.
DR SMART; SM00050; DISIN; 1.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE; PS00214; DISINTEGRIN_2; 1.
DR KEGG; Glycocalyx; Transmembrane.
DR SIGNAL; 18
FT PROPEP 19 176
FT CHAIN 177 754
FT DOMAIN 19 668
FT TRANSMEM 669 689
FT DOMAIN 690 754
FT DOMAIN 177 401
FT DOMAIN 402 488
FT DOMAIN 489 668
FT DISULFID 310 389
FT DISULFID 350 373
FT DISULFID 460 473
FT CARBOHYD 84 84
FT CARBOHYD 167 167
FT CARBOHYD 174 174
FT CARBOHYD 184 184
FT CARBOHYD 584 584
FT CARBOHYD 668 668
FT CONFLICT 369 370
FT CONFLICT 374 374
FT CONFLICT 518 518
SQ SEQUENCE 754 AA; 85582 MW; 64D04B8C7992E281 CRC64;

Query Match
Best Local Similarity 15.9%; Score 750; DB 1; Length 754;
Matches 237; Conservative 138; Mismatches 337; Indels 58; Gaps 17;

QY 1 MLQGLLPVSLLSAVSAIKE--LPQVKYEVVYPIRLHPLHREAKEPEQOE---QWET 55
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 56 ELKXKMTNGKIAVLKKNKLLAPGYTETVYNSGKEITTSQIMWVYVYQGHILAEK 115
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 58 ELLYEIKLNKRTLVLLHLLRSREFLGNSYSEFYSMKGGAFTRHPOIMDHCFYQGSIVHEY 117
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 115 VSWASISTWGLRGVYWSQGWRYTEPLSPHRCQEHALWKYX---PWEKYNWSTGWK 172
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 118 DSAASISTCNGLGFRINDQYLIPEVK--YSDEGEHLVFKYNLRVPYGANYSCT--- 171
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 173 GVLWVHNLQQNIALPATKLVKWKVQEHKYEYLVLMNGEWKRYNENQWETKRVW 232
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 172 ---ELNFRKTVPGDNESEEDSKIGHD-EKYVELFIVADTVYRRGHHPINKLRNIW 227
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 233 EMANYVNLKYLKLNTHVALVQNEITWVWKIKITPNASWTLENKSKWGSVLSRRKHWI 292
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 228 GWNVFNMIYKTNHVTLVGIEITWTHEDKELYSNIEITLLRFSFWQEKILKTRKDFH 287
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 293 AQLITATELAGTTVGLAWMSTWSP-YSGVGVQVHNSNLLR-----VACTMAHEGHNWGM 347
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 288 VLLSGKWLSSHVGISYPGMCLPYSTSIK---DLLPDTNIIANMAHQLSHNLGM 343
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 348 WHWYSWKNPSTIIVWVKALSWIPTW-WSSWSRLSYKWKWKLNLWNAFLP 406
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 344 QHDEFPCTCFSGKCVMSDGS--IPALKFKSCSQNHQVLYKDYKTCMLNIPPEYFPHD 401
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```


QY 351 WTSWKPSTIWWKALSWITPTWSSWSRLSYKWKWKLNSWLNWAPLPTWISTPIW 410
 Db 349 TKFCTCGAKPCIMFGKESIPPKFSGSYDQNKYLLKYNPKCILDPPKDKDIASPAVC 408
 QY 411 GNLQVEMGWMMWMTGTEWNTIWWAKTWKIKATWQWALGEMWKEWQWKKAGWVWRPAKW 470
 Db 409 GNEIWESECECDSPADCNPCDDATCKLKPAGCEGNGECDDCKKIRKAGTECRPARD 468
 QY 471 EHWLPKXNGKSNWPMWQVNGVPMHKGHWLWGTPLQEWTELWGPCTEVAWKS 530
 Db 469 DCDVAEHCTGQSAECPRNEFRNGQPCLNNSGYCVNGDCPIMLNQCIALFSPSATVAQDS 528
 QY 531 MYNRNKGSKYVWR---VWTLIPWKAHWKGLWQGGSW--NLFWKGRIVTWLW 585
 Db 529 CFORLQSGSYGYCTKEIGYGRFPDQVKGRGLYCLDLSFKKNCKNDY----- 582
 QY 586 KTWPEWTSQETGMVANGTKWGNKVNWNAEWWTEKAY 624
 Db 583 -----SYADENKGIPEGTCKEDGKVCINRKCVDWNTAY 616

RESULT 12
 DISF TRIFL
 ID DISF TRIFL STANDARD; PRT; 483 AA.
 AC P18619; O8:IS2;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DE Zinc metalloproteinase flavoridin precursor (EC 3.4.24.-) [Contains:
 DE Disintegrin flavoridin (RGD-containing peptide) (Platelet aggregation
 DE activation inhibitor)]
 OS Trimeresurus flavoviridis (Habu).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroides;
 CC Viperidae; Crotalinae; Trimeresurus.
 OX NCBI_TaxID=88087;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Venom gland;
 RX MEDLINE=22072855; PubMed=12076658;
 RA Kishimoto M., Takahashi T.;
 RT "Molecular cloning and sequence analysis of cDNA encoding flavoridin,
 RT a disintegrin from the venom of Trimeresurus flavoviridis.";
 RN Toxicon 40:1033-1040(2002);
 RP SEQUENCE OF 414-483.
 RC TISSUE=Venom;
 RX MEDLINE=90305100; PubMed=2364514;
 RA Musial J., Niewiarowski S., Rucinski B., Stewart G.J., Cook J.J.,
 RA Williams J.A., Edmunds L.H. Jr.;
 RT "Inhibition of platelet adhesion to surfaces of extracorporeal
 RT circuits by disintegrins. RGD-containing peptides from viper
 RT venoms.";
 RN Circulation 82:261-273(1990).
 RP [3]
 RP DISULFIDE BONDS OF DISINTEGRIN FLAVORIDIN.
 RX MEDLINE=92387379; PubMed=1516704;
 RA Calvete J.J., Wang Y., Mann K., Schaefer W., Niewiarowski S.,
 RA Stewart G.J.;
 RT "The disulfide bridge pattern of snake venom disintegrins, flavoridin
 RT and echistatin.";
 RN FEBS Lett. 309:316-320(1992).
 RP [4]
 RP DISULFIDE BONDS OF DISINTEGRIN FLAVORIDIN.
 RX MEDLINE=93360274; PubMed=8355276;
 RA Klaus W., Broger C., Gerber P., Senn H.;
 RT "Determination of the disulfide bonding pattern in proteins by local
 RT and global analysis of nuclear magnetic resonance data. Application
 RT to flavoridin.";
 RN J. Mol. Biol. 232:897-906(1993).
 RP [5]
 RP STRUCTURE BY NMR OF 414-483.

RX MEDLINE=93360275; PubMed=8355277;
 RA Senn H., Klaus W.;
 RT "The nuclear magnetic resonance solution structure of flavoridin, an
 RT antagonist of the platelet GP IIb-IIIa receptor.";
 RN J. Mol. Biol. 232:907-925(1993).
 CC -1- FUNCTION: Zinc metalloproteinase flavoridin is a zinc protease
 CC from snake venom (by similarity).
 CC -1- FUNCTION: Flavouridin inhibits fibrinogen interaction with platelet
 CC receptors expressed on glycoprotein IIb-IIIa complex. Acts by
 CC binding to the glycoprotein IIb-IIIa receptor on the platelet
 CC surface and inhibits aggregation induced by ADP, thrombin,
 CC platelet-activating factor and collagen.
 CC -1- COFACTOR: Binds one zinc ion (by similarity).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: Belongs to peptidase family M2B.
 CC -1- SIMILARITY: Contains 1 disintegrin domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR PDB; 1FVL; 29-JAN-96.
 DR InterPro; IPR001762; Disintegrin.
 DR InterPro; IPR006025; Pept_M2B.
 DR InterPro; IPR001590; Peptidase M12B.
 DR Pfam; PF00200; disintegrin; 1.
 DR PRINTS; PR00289; DISINTEGRIN.
 DR PRODOM; PD000664; Disintegrin; 1.
 DR SMART; SM00050; DISIN; 1.
 DR PROSITE; PS0215; ADAM MBRO; 1.
 DR PROSITE; PS00427; DISINTEGRIN 1; 1.
 DR PROSITE; PS0214; DISINTEGRIN 2; 1.
 DR PROSITE; PS00142; ZINC PROTEASE; 1.
 KW Hydrolase; Metalloprotease; Zinc; Blood coagulation; Platelet;
 KW Cell adhesion; Zymogen; Glycoprotein; Signal; 3D-structure.
 FT SIGNAL 1 20
 FT PROPEP 21 190
 FT CHAIN 191 395
 FT PROPEP 396 413
 FT CHAIN 414 483
 FT DOMAIN 197 395
 FT DOMAIN 403 483
 FT SITE 462 464
 FT METAL 333 333
 FT ACT_SITE 334 334
 FT METAL 337 337
 FT METAL 343 343
 FT DISULFID 417 432
 FT DISULFID 419 427
 FT DISULFID 426 449
 FT DISULFID 440 446
 FT DISULFID 445 470
 FT DISULFID 458 477
 FT CARBOHYD 263 263
 FT CARBOHYD 293 293
 FT CONFLICT 455 456
 FT TURN 425 426
 FT STRAND 427 428
 FT TURN 429 432
 FT STRAND 433 434
 FT TURN 436 437
 FT STRAND 446 446
 FT STRAND 451 451
 FT STRAND 457 459
 FT STRAND 468 469
 SQ SEQUENCE 483 AA; 54513 MW; 3B943C81B6C7B1C3 CRC64;
 Query Match 12.2%; Score 574; DB 1; Length 483;

```

venom metalloproteinase, atrolysin C (form d).";
-!- FUNCTION: This protein is a zinc protease from snake venom that acts in hemorrhage. It cleaves type IV collagen and gelatin.
-!- CATALYTIC ACTIVITY: Cleavage of 5-His--|-Leu-6, 10-His--|-Leu-11, 14-Ala--|-Leu-15, 16-Tyr--|-Leu-17 and 23-Gly--|-Phe-24 of insulin B chain. With small molecule substrates prefers hydrophobic residue at P2. With small residue such as Ala, Gly at P1.
-!- COFACTOR: Binds 1 zinc ion and 1 calcium ion per subunit.
-!- MISCELLANEOUS: The sequence shown is that of HT-D.
-!- SIMILARITY: Belongs to peptidase family M12B.
-!- SIMILARITY: Contains 1 disintegrin domain.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch.
-----
EMBL; U01237; AAA03352.1; -.
FIR; S41610; HYRSAC.
FDB; 1ATL; 15-OCT-95.
FDB; 1DTH; 12-FEB-97.
FDB; 1HTD; 15-SEP-95.
DR MEROPS; M12.144; -.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001590; Peptidase_M12B.
DR InterPro; IPR002870; Peptidase_M12B_N.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR PROSITE; PS00215; ADAM_MEPRO; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Collagen degradation; Hydrolase; Metalloprotease; Calcium-binding;
KW Metal-binding; Zinc; Signal; Pyroliidone carboxylic acid;
3D-structure.
FT SIGNAL; 1 20
FT PROPEP 21 190
FT CHAIN 191 393
FT CHAIN 394 414
FT WOD RES 191 191
FT METAL 333 333
FT ACT SITE 334 334
FT METAL 337 337
FT METAL 343 343
FT METAL 343 343
FT DISULFID 308 388
FT VARIANT 348 355
FT VARIANT 191 191
FT VARIANT 371 371
FT STRAND 197 205
FT HELIX 207 212
FT TURN 213 215
FT HELIX 217 235
FT HELIX 236 238
FT TURN 239 239
FT STRAND 240 249
FT HELIX 262 275
FT TURN 276 276
FT HELIX 277 280
FT STRAND 285 290
FT STRAND 300 301
FT TURN 304 305
FT TURN 307 308
FT TURN 310 312
FT STRAND 314 318
FT HELIX 324 338
FT TURN 339 340
FT TURN 346 347
FT STRAND 349 350
FT TURN 351 352
FT STRAND 353 355
FT TURN 356 357
FT POTENTIAL.
FT HEMORRHAGIC METALLOPROTEINASE HT-D.
FT DISINTEGRIN (POTENTIAL).
FT PYRROLIDONE CARBOXYLIC ACID.
FT ZINC (CATALYTIC).
FT ZINC (CATALYTIC).
FT ZINC (CATALYTIC).
FT ZINC (CATALYTIC).
FT MISSING (IN SOME CHAINS).
FT D -> A (IN HT-C).

```


KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
 FT Tracsmembrane; EGF-like domain; SH3-binding; Alternative splicing.
 FT SIGNAL 1 25
 FT PROPEP 26 203
 FT CHAIN 204 956
 FT DOMAIN 204 700
 FT TRANSMEM 701 721
 FT DOMAIN 722 956
 FT DOMAIN 204 410
 FT DOMAIN 416 502
 FT DOMAIN 435 438
 FT DOMAIN 503 650
 FT DOMAIN 651 683
 FT SITE 834 840
 FT SITE 839 845
 FT SITE 133 133
 FT METAL 346 346
 FT ACT SITE 347 347
 FT METAL 350 350
 FT METAL 356 356
 FT DISULFID 321 404
 FT DISULFID 361 388
 FT DISULFID 475 482
 FT DISULFID 655 665
 FT D-SULFID 659 671
 FT D-SULFID 673 682
 FT CARBOHYD 145 145
 FT CARBOHYD 445 445
 FT CARBOHYD 448 448
 FT CARBOHYD 646 646
 FT VARSPLIC 903 956

Query Match
 Best Local Similarity 11.4%; Score 539.5; DB 1; Length 956;
 Matches 215; Conservative 114; Mismatches 345; Indels 153; Gaps 25;

36 LRPLHKKRAKEP-----QQR-----QWETE-----LKYKMTI--NGKI 67
 17 LQFLPRAAREPGWTSKSGSESPKQLQHELIIPQWKTSPVREKPEPLKALRWMAEGRE 76
 66 AVLYLKONKHLAPGYTETTYNSTGKEITTSQIMWVYQGHLLNEKYSWASISTWGL 127
 77 LILDLEKNEQLPAPSYTETHYSSNGEQTTTRKLEDHCFYHGTVRETELSSVTLSTCRGI 136
 128 EGYNSQGHQ-RYWIEPL-----SPIHRWGQEHALWKNPWEKNYMWSTWG 170
 137 EGLITVSSNLSYVIEPLDPSKQHLIYRSEHLKPPPGNGCFEHSKPTTDM----- 187
 171 MWGLWAEWLQNIAPLPAKLVKLRKQVQEH-----KYIEYVLWNGEKRYNENQWE 226
 188 -----ALQFTQTKGPRMKREDLNSMKYVELYLVADYLEFQKRRDQDA 233
 227 IRKRWEMANYNMLEYKKNITVALVGMELITWKKIKITNASWLENKSKWRSVLRSR 286
 234 TKHKLIEIANYVDRYSRINRIALVGLVETHGNCNCRSENPSYTLSSFLSWRKLIA- 292
 287 KRWIAQLITATELAGTVGLAMWSTMWSPY-SVGYYQVHWSNLLRVLVAGTMAHEMGHW 345
 293 QKYNHNAQLITGMSFHTTIGLPLMACSYVQSGVNMDSHNAIGVAAATMAHEMGHNF 352
 346 GMMHWYWSKWPSTI---WVWKKALSWIPTWWSWSRLSYKWKWMLSNWLNAPLPT 402
 353 GYTHDSADCCSAAADGGCNAATGHPFPKVFNGCNRRELDRLVQSGGGLSNMPPDR 412
 403 WIISTPIWGNQVMEGEMWNGTSEETNIWAKTWKIKATQWALGEWWEKQWTKAG 462

413 MEYGRRCNGYLEDEGEBCDCEEEECNNPCNNASCTLRPGAECAGHSCCHCKLLAPG 472
 463 MYWRPAKWEWLPPEMNGKSGNPPWVWQVNGWPHHGHKWLMTGTPTLOEQWTLWGP 522
 473 TLRCSARQCDELPECTGKSPHCPNPNFYOMDGTCEGGOAYCNGMCLTYQEQCQLWGP 532
 523 GTEVAKWGWYNR-NEGGSKYGYRRVWMTLIPKAWN--TWGKLWQGG----- 569
 533 GARPAPDLCEBKVNVAAGDTFGNGKV-MNGEHRKCNMRDAKCGKICQCSSEARPLESNV 591
 570 --SWNLPMKGRIVTWLTKTW-MPEWTQEI--GMVANGTKWGMKWKVWNAEWVLEKAY 624
 592 PIDITIMNGROIQCRGTHVYRGPEEGDMLNPLGLVMTGTCGYNHCIFEGQ-CRNTSFP 650
 625 KSTNWSKWKGHAVWHELOWMEBGMIPPPWWWSSVWVHNS-----IVVG 670
 651 ETEGGCKNGHGVNCCNQNCHCLPGWAPP--FNTFGHGSIDSIDGMPPEPVVVG 707
 671 VLNPMVTVVVMVIRHQSREKQKQRPSTLTGTTRPHQ----- 712
 708 VL---VAIVLVLMLMYCCRNKQLQKPSALPSKLRQCFSPFVVSQNSGTGHANP 764
 713 -----KEK---PQWKAQVQPEKSMQKPHVWLPVGEPPA 746
 765 TFKLQTEGKREKRVINTPEILR--KPSQPPPRPEEDY---LAGGSPPA 806

RESULT 15
 HRTE CROAT
 ID HRTE CROAT STANDARD; PRT; 478 AA.
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hemorrhagic metalloproteinase Ht-E precursor (EC 3.4.24.44)
 DE (Atrolysin B) (Hemorrhagic toxin E).
 OS Crotalus atrox (Western diamondback rattlesnake).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Lepidodonta; Squamata; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Viperidae; Crotalinae; Scleroglossa; Serpentes; Colubroidea;
 OX NCBI_TaxID=8730;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Venom;
 RA Hite L.A., Shannon J.D., Bjarnason J.B., Fox J.W.;
 "Sequence of a cDNA clone encoding the zinc metalloproteinase
 RT hemorrhagic toxin e from Crotalus atrox: evidence for signal,
 RT zymogen, and disintegrin-like structures.";
 RL Biochemistry 31:6203-6211(1992).
 CC -!- FUNCTION: This protein is a zinc protease from snake venom that
 CC acts in hemorrhage.
 CC -!- FUNCTION: Inhibits fibrinogen interaction with platelet receptors
 CC expressed on glycoprotein IIb-IIIa complex. Acts by binding to the
 CC glycoprotein IIb-IIIa receptor on the platelet surface and
 CC inhibits aggregation induced by ADP, thrombin, platelet-activating
 CC factor and collagen (By similarity).
 CC -!- CATALYTIC ACTIVITY: Cleavage of 3-Asn-[Gln-4, 9-Ser-1-His-10 and
 CC 14-Ala-Leu-15 bonds in insulin B chain and 14-Tyr-[Gln-15 and 8-
 CC Ala]-Ser-9 in A chain. Cleaves type IV collagen at 258-Ala-[Gln-
 CC 259 in alpha-1-(IV) and at 191-Gly-[Leu-192 in alpha-2-(IV)].
 CC -!- COFACTOR: Binds 1 zinc ion per subunit (probable).
 CC -!- SIMILARITY: Belongs to peptidase family M12B.
 CC -!- SIMILARITY: Contains 1 disintegrin domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

Search completed: June 9, 2004, 18:01:15
Job time : 20 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2004, 17:57:25 ; Search time 52 Seconds
(without alignments)
4702.435 Million cell updates/sec

Title: US-10-726-148a-15
Perfect score: 4712
Sequence: 1 MQGLPVSLLSVAISAIAK.....PPTVKNPMSPTKSNPKA 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 5

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacterioph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	%	Query	Description
No.	Score	Match	Length	ID	Description
1	1812	38.5	718	11 Q7TSB4	Q7teb4 rattus norv
2	751	15.9	612	13 Q8UVG0	Q8uvu0 bothrops er
3	740.5	15.7	620	13 Q42138	Q42138 agkistrodon
4	733.5	15.6	609	13 Q8JGN1	Q8jgn1 najia mossam
5	718.5	15.2	609	13 Q8JIR2	Q8jir2 trimeresuru
6	702	14.9	470	4 Q15204	Q15204 homo sapien
7	699	14.8	610	13 Q9YI20	Q9yi20 agkistrodon
8	698.5	14.8	600	13 Q9PVK7	Q9pvk7 najia najia
9	695.5	14.8	617	13 Q90499	Q90499 echis pyram
10	693.5	14.7	609	13 Q90282	Q90282 crotalus at
11	693.5	14.7	614	13 Q8JIR1	Q8jir1 trimeresuru
12	689	14.6	606	13 Q98UF9	Q98uf9 bothrops ja
13	684.5	14.5	610	13 Q93523	Q93523 bothrops ja
14	678	14.4	610	13 Q8Q388	Q8qg88 bothrops in
15	675	14.3	612	13 Q7T046	Q7t046 vipera lebe
16	673	14.3	825	11 Q8R3D3	Q8r3d3 mus musculu

17	666.5	14.1	744	11 Q8C269	Q8c269 mus musculu
18	665.5	14.1	612	13 Q90Z13	Q90z13 trimeresuru
19	654	13.9	610	13 Q9DGB9	Q9dgb9 crotalus at
20	652.5	13.8	609	13 Q9W6M5	Q9w6m5 agkistrodon
21	648	13.8	610	13 Q8AW15	Q8aw15 agkistrodon
22	640	13.6	604	13 Q9PT48	Q9pt48 atractaspis
23	600	12.7	478	13 Q7SZD9	Q7szd9 agkistrodon
24	595	12.6	476	13 Q9YI19	Q9yi19 agkistrodon
25	593.5	12.6	483	13 Q805F4	Q805f4 agkistrodon
26	588.5	12.5	483	13 Q805F6	Q805f6 agkistrodon
27	586.5	12.4	483	13 Q91AB0	Q91ab0 agkistrodon
28	586	12.4	697	13 Q7ZYZ9	Q7zyz9 brachydanio
29	585.5	12.4	409	13 Q8QG89	Q8qg89 bothrops in
30	576.5	12.2	479	13 Q9PWJ0	Q9pwj0 agkistrodon
31	576.5	12.2	505	13 Q73795	Q73795 agkistrodon
32	573	12.2	405	13 Q7SZD8	Q7szd8 agkistrodon
33	573	12.2	483	13 Q7SZE0	Q7sze0 glydius sa
34	570	12.1	482	13 Q9PVK9	Q9pvk9 agkistrodon
35	566	12.0	466	13 Q91AX6	Q91ax6 agkistrodon
36	566	12.0	466	13 Q91AX7	Q91ax7 agkistrodon
37	566	12.0	477	13 Q98SP2	Q98sp2 bothrops ja
38	560.5	11.9	479	13 Q9PW78	Q9pw78 agkistrodon
39	559.5	11.9	481	13 Q90YA6	Q90ya6 trimeresuru
40	552.5	11.7	549	13 Q90500	Q90500 echis pyram
41	551	11.7	407	13 Q92032	Q92032 agkistrodon
42	549	11.7	478	13 Q90W25	Q90w25 trimeresuru
43	549	11.7	481	13 Q90YA7	Q90ya7 trimeresuru
44	543	11.5	411	13 Q92031	Q92031 agkistrodon
45	542.5	11.5	414	13 Q8CJ50	Q8cj50 crotalus mo

ALIGNMENTS

RESULT 1

Q7TSB4 PRELIMINARY; PRT; 718 AA.

ID Q7TSB4
AC Q7TSB4;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DE ADAM28 isoform-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Epididymis;
RA Oh J., Cho C.;
RT *Cloning and characterization of epididymal ADAMs.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY283187; AAP56236.1; -.
SQ SEQUENCE 718 AA; 80616 MW; E1CBA1067FA904CE CRC64;

Query Match	38.5%;	Score 1812;	DB 11;	Length 718;
Best Local Similarity	60.0%;	Pred. No. 7.6e-126;		
Matches 391;	Conservative 84;	Mismatches 175;	Indels 2;	Gaps 2;
QY	5	LLPVSLLS-VAVSAIKELPGVKYEVVYPIRLHPLHKREAKEPEQOQOWETELKYMTI	63	
Db	6	LLVWFLLSPVASAIKELPKAKYEVVYPIRLHLLHKREKPEKPEKPEKPEKPEKPEKPEK	65	
QY	64	NGKIAVLKONKLLAPGYTETYNSTGKETTTSQIMWVYQGHILNEKYSNASIST	123	
Db	66	NGKVVLYLKKONKLLAPGYLETYNSSGNKVTTSQIMDSYQGHILNEKDSASISM	125	
QY	124	WRGLGVWQWQRYWIEPLSPHFWGQEHAKYKPNWKNYKYNWSTWGMVLAHMLQQN	183	
Db	126	CQGLRGYFSQADRYEIEPLSLDEQAHALFKDPKEDQGNCSGCVDDALMLQGLHQD	185	
QY	184	IALPATKLVKLRKVKQEHKEKYLEYLVLVNKGWKYKYNENQWIRKRVWEMANYVNNLYK	243	

```
Db 186 VVLPAIRLILKNDGKQVQPKXKIEYVVVLDNGBEFKKYNKNLDEIKRKHVHEMANYVMYLN 245
Qy 244 KLNTEVALVGMEIWTWKWKIKITPNASWTLENWSKRGSVLSRRKRHWIAQILITATELAG 303
Db 246 KLGARVALVGMEIWTDEDKIKITPDANTLENFSKRGNDLLKRKHHDVAQLISSTDFSG 305
Qy 304 TTUGLAKYSTWSPY-SVGVVQWHSNLLRVAGTVAHMGHNWGMHWYWKWPFSTIWW 362
Db 306 STVGLAFSSSCSPYSHVGIQDHSNHLRVAGTVAHMGHNWGMHWYWKWPFSTIWW 365
Qy 363 YKALSWIPTWWSWSRLSYKWKWEMKLSNWLKAPLPTLIISPIGNOLVEMGEWMM 422
Db 366 MEQSLRHPMTDFSCSDNTRFLEKLSHCLFNSPLPSDLSIPVCGNQLENNEBCD 425
Qy 423 WGTSEENTNIWWAKTWIKATWQWALGEMWKKWKKAGMYWRPAKWEMWMLPEMWNCKS 482
Db 426 CGTPKECTNKCDAETCKIKAGFQCALGECCEKCKLKKPGVVCRAAKDECDLPEMWCCKS 485
Qy 483 GNWPKWQVNGWPHHGHGHLWLTGTPLOEOWTELGPGTEVANKSWYRNNEGSKYG 542
Db 486 SHCPVDFRVRNGFPQNGHGYCLKNCPTLQQQCMVMWGPETKVAKSCYKQNEGSKYG 545
Qy 543 YWRRVWMTLIPWKANWTWGLWMOGSGWMLPWKGRIVTWLTWKTWMPENTSOEIGMVAN 602
Db 546 YCHVENGTHMCKAKDAMCGKLFCEGSSDLPWKGLTIAFLTKLFDPEIDINQGYDMVAN 605
Qy 603 GTKWGNKWINAEWWTETKAYKSTWSSKKGHGAHVWHEELQWBERGMIPP 654
Db 606 GTKGNKVKCINAEACADMEKTKYKSAKCKGKHAHVCDHELCQCKEGWAPP 657

RESULT 2
Q8UVG0 PRELIMINARY; PRT; 612 AA.
AC Q8UVG0;
AD Q8UVG0;
DT 01-MAR-2002 (TremBLrel. 20, Created);
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update);
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update);
DE Berythraceivase.
GN ERV1.
OS Bothrops erythromelas (Caatinga lance head).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC viperidae; Crotalinae; Bothrops.
OX NCBI_taxid=44710;
RN [1]
RP SEQUENCE FROM N.A.
RA Silva M.B., Schattner R., Ramos C.R., Lazzari M.A., Azevedo I.L.M.J.,
RA Guarnieri M.C., Pozner R.G., Ho P.L., Tavassi A.M.C.;
RA "A prothrombin activator from the Bothrops erythromelas snake venom:
RT characterization and molecular cloning.";
AL Submitted (NOV-2001) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF450503; AAL47169.1; -.
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
DR GO; GO:0008270; P:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR004586; ADAM_cysteine.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR001590; Peptidase M12B.
DR InterPro; IPR002870; Peptidase M12B_N.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF00200; disintegrin.1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR PRINTS; PR00289; DISINTEGRIN.
DR ProDom; PDO00664; Disintegrin; 1.
DR SMART; SM00608; ACR; 1.
DR SMART; SM00050; DISIN; 1.
DR PROSITE; PS02015; ADAM_MPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; 1.
DR PROSITE; PS02014; DISINTEGRIN_2; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 612 AA; 68531 MW; C04E8FAFF983F606 CRC64;
```

```
Query Match 15.9%; Score 751; DB 13; Length 612;
Best Local Similarity 34.3%; Pred. No. 2,1e-47;
Matches 217; Conservative 113; Mismatches 274; Indels 28; Gaps 11;

Qy 1 MLOGLLPVSLLSV---AVSAIKELPGVKYKVVVPIRLPLHKKRAKPEQEQWETEL 57
Db 1 MIQVLLVITICLAEFPYQGSIILESNGVNDYEVVYPRKVLTALSK-GAVHP----KYEDAM 55
Qy 58 KYKMTINKIAVLKKNKLLAPGYTETYNSTKEITTSPIQIMWWTYQGHILNEKVS 117
Db 56 QYEFKVNGBPVVHLKKNKGLFSEDYSEIHYSFGREITTYPLVEDHCYHGRIONADS 115
Qy 118 WASISTWRGLRGYWGQWORYWIBLSPILHRGQBHALMKYNPNPKYNYSTWGMGVLA 177
Db 116 SASISACNGLKHGHPQLQEMYLIEPFLPD--SEAHVFKYENVEKEDEAP-KMGCVTET 172
Qy 178 HWLQQNIALPATKLVKL--KMKVQEHKYEIVYLVLNNGEWKRYNENQWEIKRWEMA 235
Db 173 NW-ESDEPIKASLNLNLTPEQQAYLDKAKYVEFVVLDHGMTKYKADDDLKIRIYEV 231
Qy 236 NYVNNLYKLANTHVALVGMETWTKWKIKITPNASWTLENWSKRGSVLSRRKRHWIAQL 295
Db 232 NTMNEMFIPLNLCVALTGLEIWSKDKINVTSESWEFLILFTNRGADLLKRSKSHDNAQL 291
Qy 296 ITATBLAGTTCGLAWMTWMSPY-SVGVVQWHSNLLRVAGTVAHMGHNWGMHWYWK 354
Db 292 LTNTDFDGSITIGRAHIGSMCHPYLSVGIQDYSFVNLVASTVAHMGHNLMGHNDTC 351
Qy 355 KWPESTIWWKALSWIPTWSSWSRLSYKWKWEMKLSNWLKAPLPTLIISPIGNOL 414
Db 352 TCGAPSCVMAAAISKDPSKLSNCSQEVQRYKLIKNRPQCLNKLPLRTDILSPVCGNEL 411
Qy 415 VEMGEWMMWGTSEENTNIWWAKTWIKATWQWALGEMWKKWKKAGMYWRPAKWEMWML 474
Db 412 LEVGECDGCTPENCRCPCCNATTCKLTPGSCQVEGLCCDQCRFRKGTGECRAAKHCDL 471
Qy 475 PEMWNGKSGNPPWQWVNGWPHHGHGHLWLTGTPLOEOWTELGPGTEVANKSWYRN 534
Db 472 PESCTGSADCFMDDFORNGHPQNNNGCYVNGKCPTMENQCIDILVGRATVAEDSCFKD 531
Qy 535 NEGSKYGYWRRVWMTLIPWKANWTWGLWMOGSG--WNLFPWKGRIVTWLTWKTWMPBW 592
Db 532 NQKNGDYGYCKENCKKIPCPSPQDVKCGRLYCNDSPGQNPCK-----CIYFPR- 581
Qy 593 TSQEIQGVANGTKWGNKWINAEWWTETKAY 624
Db 582 -NEDRGVLPQTKCADGKVCNRRHCVDAATAY 612

RESULT 3
Q42138 PRELIMINARY; PRT; 620 AA.
AC Q42138;
AD Q42138;
DT 01-JAN-1998 (TremBLrel. 05, Created);
DT 01-JAN-1998 (TremBLrel. 05, Last sequence update);
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update);
DE Metalloproteinase-disintegrin-like protein (EC 3.4.24.1).
OS Agkistrodon contortrix laticinctus (Broad-banded copperhead);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Agkistrodon.
OX NCBI_taxid=37195;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=95314311; PubMed=7793974;
RA Selistre de Araujo H.S., Ownby C.L.;
RT "Molecular cloning and sequence analysis of cDNAs for
RT metalloproteinases from broad-banded copperhead Agkistrodon contortrix
RL laticinctus.";
RL Arch. Biochem. Biophys. 320:141-148(1995).
RN [2]
RP SEQUENCE FROM N.A.
```

```

RX MEDLINE=98052412; PubMed=9392519;
RA Seliestre de Araujo H.S., de Souza D.H., Omby C.L.;
RT "Analysis of a cDNA sequence encoding a novel member of the snake
RT venom metalloproteinase, disintegrin-like, cysteine-rich (MDC) protein
RT family from Agkistrodon contortrix latinctus."
RL Biochim. Biophys. Acta 1342:109-115(1997).
DR EMBL; U86634; AAC18911.1; -
DR HSSP; P18619; 1FVL.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
DR GO; GO:0038270; F:zinc ion binding; IEA.
DR GO; GO:0037229; P:intracellular-mediated signaling pathway; IEA.
DR GO; GO:0036508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR006386; ADAM cysteine.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR001590; Peptidase M12B.
DR InterPro; IPR002870; Peptidase M12B.
DR InterPro; IPR006025; Pept M, Zn_BS.
DR Pfam; PF00200; disintegrin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR PRINTS; PR02289; DISINTEGRIN.
DR ProDom; PD00664; Disintegrin; 1.
DR SMART; SM00608; ACR; 1.
DR SMART; SM00050; DISIN; 1.
DR PROSITE; PS0215; ADAM_MPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; 1.
DR PROSITE; PS0214; DISINTEGRIN_2; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Integrin.
SQ SEQUENCE 620 AA; 69512 MW; 9016AFEB5AE0B387 CRC64;

Query Match 15.7%; Score 740.5; DB 13; Length 620;
Best Local Similarity 34.8%; Pred. No. 1.3e-46;
Matches 221; Conservative 110; Mismatches 279; Indels 31; Gaps 12;

QY 1 MLQGLLPVSLLSV-----AVSAIKELPGVKYEVVYPIRLHLHREAKEPEQOEOMETE 56
Db 1 MIQVLL-VTLCLAVFPYQSGSIILSGVNDYEVVYPRKTVLPKAV-----QPKYEDA 54
QY 57 LKYMTINGKIAVLKKNKLLAPGYTETYNSTKEITTSQIMMWWYQGHILNEKV 116
Db 55 MQYEFKVGCEPVVHLKXNQLFSKDYSETHYSPDGREITTYPLVEDHCYHGHRIENDAD 114
QY 117 SWASISTWRLGRYWGQWQRYWIEPLSPHRRGOEHALWKYNPWEKYNWSTWGMGVLM 176
Db 115 STASISACNGLGHFLQGEWYLDLKLDP--SEAHAVFKYENVEKEDAP-KMGCVT- 170
QY 177 AHWLQONIALPATKLVKLKWRKVQ-----BHEKYEIYVLVWNGEWKRYNENOWEIRKRW 232
Db 171 QWSEYE--PIKASQNLNTPQOAYLDAKKVFEVVLHDGMVTKYKNDLKIIRIP 227
QY 233 EMANYVMYKXKLNTHVALVGMIEWTWKIKITPNSWTLNWKSKWGVLSRRKRHWI 292
Db 228 EIVTWNMFIPLNIRVALICLEIWSDKFNNTSAAVNTSISFRNWRATDLKLRKSHON 287
QY 293 AOLITATELAGTGVGLAWSTWSP-YSGVQWQWHLNLRVAGTWAHEGHNWCKWQHW 351
Db 288 AOLITVIDFDGPTIGKAYASMCDPKRSW::IQDSTINLMVAVTWAHEGHNWGLZMDHDE 347
QY 352 YSWKWPSTIWWKALSWIPTWSSWSRLSYKWKWHLNLMNAPITWTIISTPIWG 411
Db 348 KYCTCGAKSCVMKALSRQPSKLFNSCQEDYRKYLKRRPKILNPNNGTIVSPVCG 407
QY 412 NQIVGEVWWTGTSBENTNIWWAKTIKATQWALGEWKEWKKWKGAGVWRPAKWE 471
Db 408 NELLEVEGECDGSPNCPNCPDAAATCLTPGSCQADGVCCDQCRCFRAGTECRQXOD 467
QY 472 WMLPEMNGSKGNWPNWRQVNCWPHHKGHWLMGTWPTLQEWTELCGPGTEVAKWS 531
Db 468 CDMADLCQGSACPTDRFQRNGHPCINDNGYCYNRTCPTLKNQCITYFPQPNAAVAKDC 527
QY 532 YNRNEGSKYGYWRVWVWTLIPKANWTVWGLWQGSNLPWKGRIVTWTWKTWPE 591

528 FKNGKSNHNYCRKNGKKIPCAPODIKGRLYC---FRLPKKKICSVI-----YTP- 579
592 WTSQETGMVANGYKGNKKNWVNAWVWIEKAYKSTWSSK 632
580 -TDEDIGMVLPGTKCEDGKVCNGHCVDNVIAVYKSTTGFSSQ 619

RESULT 4
Q8JGN1 PRELIMINARY; PRT; 609 AA.
AC Q8JGN1;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mocarhagin 1.
OS Naja mossambica mossambica.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Elapinae; Naja.
OX NCBI_TaxID=196380;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RA Sako D., Shaw G.D.;
RT "Molecular characterization of mocarhagins: A multi-gene family of
RT metalloproteinases expressed in cobra venom."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY101383; AAM51550.1; -
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR006586; ADAM_cysteine.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR001590; Peptidase M12B.
DR InterPro; IPR002870; Peptidase M12B.
DR InterPro; IPR000169; SHprot acsite.
DR Pfam; PF00200; disintegrin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR PRINTS; PR02289; DISINTEGRIN.
DR ProDom; PD00664; Disintegrin; 1.
DR SMART; SM00608; ACR; 1.
DR SMART; SM00050; DISIN; 1.
DR PROSITE; PS0215; ADAM_MPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; 1.
DR PROSITE; PS00139; THIOI_PROTEASE_CYS; 1.
SQ SEQUENCE 609 AA; 68176 MW; FE717DCAE344A40D CRC64;

Query Match 15.6%; Score 733.5; DB 13; Length 609;
Best Local Similarity 34.8%; Pred. No. 4.2e-46;
Matches 214; Conservative 104; Mismatches 277; Indels 35; Gaps 12;

QY 1 MLQGLLPVSLLSV-----AVSAIKELPGVKYEVVYPIRLHLHREAKEPEQOEOMETE 56
Db 1 MIQALL-VAICLAVFPYQSGSIILSGVNDYEVVYPRKVPALSKGVQNPQPKYEDT 59
QY 57 LKYMTINGKIAVLKKNKLLAPGYTETYNSTKEITTSQIMMWWYQGHILNEKV 116
Db 60 MQYEFKVGCEPVVHLKXNQLFSKDYSETHYSPDGREITTSVPQDHCYHGHQHEAD 119
QY 117 SWASISTWRLGRYWGQWQRYWIEPLSPHRRGOEHALWKYNPWEKYNWSTWGMGVLM 176
Db 120 SSASVTSACDLGKHGPHQGETYFIEPLESD--SEAHAYKDNVEEER-EIPKICGVQ 176
QY 177 AHWLQONIALPATKLVKLKWRKVQEH-----EKYEIYVLVWNGEWKRYNENOWEIRKRW 232
Db 177 TTWESDE--PIKSSQLNTPQDRYLQAKKIEFVVVDVNMVYKTKLHVITSRVY 233
QY 233 EMANYVMYKXKLNTHVALVGMIEWTWKIKITPNSWTLNWKSKWGVLSRRKRHWI 292
Db 234 EIVTWNMFIPLNIRVALICLEIWSNGEINVSQDVAOTDLDFGEWRNKLIPRKENDN 293
```



```

370 YIPTWWSWLSLWYKWKLSNWLNAPLP*LIISTPWGNQNLVGEWMMWMTGSEEM 429
362 Y---OFSSCSVREHQRYLLRDRPQCILNKPJSTETIVSPICNGYVEVEECCDCSPADC 418
430 TNIWWAKTKIKATWQWALGEMWKEKQWKKAGVVRPAKWEMWLPWMWNGSKGNWPMWR 489
419 QSACCNATTKLQHEAQDSEECCKCKFKGAGACRAAKDDCDLPELCTGQSABCPDGV 478
490 WQVNGWPHWGHGHWLWMTWLTQWMTLPGTGVAVKWSYNEGSKYGYWVRVW 549
479 FQNGPLQCN-NGYCYNGKCPITNQICIALRGPVGVKVRSDSCFTLNQTRGGLCRMEYG 537
550 TLIIPKANWMTWGLWQGSWNL*PWKGRIVTWLTWKTWBPWTSQELGMVANGTKWGN 609
538 RKIFCAAKDVKCKGLFCKK-----RNSMTICNCSISPRDPSY-----GMVEPGTKCGDG 585
610 KVINAEVWVIEKAY 624
586 MVCNSNRQCVOKTAY 600

RESULT 9
Q90499 PRELIMINARY; PRT; 617 AA.
AC Q90499;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Metalloprotease.
GN ECHI.
OS Echis pyramidum (carpet viper).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Viperinae; Echis.
OX NCBI_TaxID=8700;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RX MEDLINE=95010025; PubMed=7925363;
RA Paine M.J.I., Moura-Da-Silva A.M., Theakston R.D.G., Crampton J.M.;
RT "Cloning of metalloprotease genes in the carpet viper (Echis pyramidum
leakeyi): Further members of the metalloprotease/disintegrin gene
family.";
RL Eur. J. Biochem. 224:483-488(1994).
DR EMBL; X78970; CRA55565.1; -.
DR PIR; S48160; S48160.
DR HSSP; P18619; 1FVL.
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR006586; ADAM_cysteine.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR001590; Peptidase M12B.
DR InterPro; IPR002870; Peptidase M12B_N.
DR Pfam; PF00200; disintegrin_1_BS.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR PRINTS; PR00289; DISINTEGRIN.
DR ProDom; PD000664; Disintegrin; 1.
DR SMART; SM00608; ACR; 1.
DR SMART; SM00050; DISIN; 1.
DR PROSITE; PS0215; ADAM_MEPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; 1.
DR PROSITE; PS0214; DISINTEGRIN_2; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Protease.
FT CHAIN 193 617 METALLOPROTEASE.
SQ SEQUENCE 617 AA; 69310 MW; 83DC3DA5F4F8A CRC64;

Query Match 14.8%; Score 695.5; DB 13; Length 617;
Best Local Similarity 35.2%; Pred. No. 2.8e-43;
Matches 193; Conservative 94; Mismatches 244; Indels 17; Gaps 10;

```

QY 1 MLOGLPVSLLLSV---AVSAIKELPGVKYKVVVPIRLHLPHKREAKEPEQOQWETEL 57
Db 1 MIQELLVWTCVAVFPYQSSIILESNGVNDYBIVFPQKNATLTKEAIOQPEQ--KYEDTM 58
QY 58 KYKMTINGKIAVLKLNKKNLLAPGYTYTNSGTKEITTSPOIMWVYQGHILNEKVS 117
Db 59 RYEFKNGEPVVLHKKNGKGFSEDIYSTHSPDQREITTPVEDHCHYHGRIONDADS 118
QY 118 WASISTWRGLGYMSQGWRYWIEPLSPIHRWGQEHALWKNPWBKNYWSWTGMGVLWA 177
Db 119 TASNSACNGLAGYMLRGETYLIEPLKIPD--SEAHAVKYENVEKEDEAP-KMGGVQT 175
QY 178 HLOQNTALPATKLVKLKRWKQOEHEKYLEYVLVWNGEWKRYNENQHEIRKRWEMANY 237
Db 176 NWSDELK-KASQVATS--EQRSYKXKIEFVVVADYIMYRKYNDSTAVRRRIEIVNI 233
QY 238 VNMLYKLNTHVALVGMETWTKWIKITPNASWTLNWSKWRGSLRRKRHWIAQLIT 297
Db 234 LNWYIVENIHVALVHIEIMSTRODITVQSAADVTLDLEGDWRKALLTRKEDHNAQLFT 293
QY 298 ATELAGTTVGLAWMTMSP--YSVGWQVMSNLLRVAGTMAHEMGNKXWYWHYWSKW 356
Db 294 GINLNGQTLGIARMGCMSPNSGVVIQYCKNLLVAITMAHELGNLXMDHDNGCNC 353
QY 357 PSTIWMKALSWIPTW--WSWSRLSYKWKWEKLSNWLWNAPLPTWILSTPIWGNQIV 415
Db 354 PDTSCIM-SAVAGPEPVFSFNCSDRYSRFNSDQSCIDNKLKTDIVSPVCGNVFV 412
QY 416 EMGSMWMTGTSEETNIWWAKTKIKATWQWALGEWKEKQWKKAGWVRPAKWEMWLP 475
Db 413 EVGECDCGSRITYCRNPNCTATCKLTQSCQADGECNOCQFPRARTECRKIDDCVP 472
QY 476 EMWNGKSNWPMRWQVNGVPMWGHKGHLMTWPTLOEQWTELMGP-----GTEVAKSW 531
Db 473 BYCTGQCECLDVPQRNGQYQSNNGCYNGNCPILKNQCIELMKPAPPAGVAVPDVC 532
QY 532 YNRNEGGS 539
Db 533 FEDNGT 540

RESULT 10
ID Q90282 PRELIMINARY; PRT; 609 AA.
AC Q90282;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DB Crotocollastatin precursor.
OS Crotalus atrox (Western diamondback rattlesnake).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodactylia; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Crotalus.
OX NCBI_TaxID=8730;
RN [1]
RP SEQUENCE FROM N.A.
RT TISSUE=venom gland;
RX MEDLINE=95251603; PubMed=7733877;
RA Zhou Q., Smith J.B., Grossman M.H.;
RT "Molecular cloning and expression of crotocollastatin, a snake-venom
protein from Crotalus atrox (western diamondback rattlesnake) which
inhibits platelet adhesion to collagen."
RL Biochem. J. 307:411-417(1995).
DR EMBL; U21003; AAC59672.1; -
DR PIR; S55270; S55270.
DR HSSP; P17494; 1XST.
DR GO; GO:0004222; E:metalloendopeptidase activity; IEA.
DR GO; GO:0008270; E:zinc ion binding; IEA.
DR GO; GO:0006508; E:proteolysis and peptidolysis; IEA.
DR InterPro; IPR006586; ADAM cysteine.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR001590; Peptidase_M12B.

DR InterPro; IPR002870; Peptidase_M12B_N.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF00200; Disintegrin; 1.
DR Pfam; PF01562; Pept_M12B_propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR PRINTS; PR00289; Disintegrin.
DR ProDom; PD000664; Disintegrin; 1.
DR SMART; SM00608; ACR; 1.
DR SMART; SM00050; DISIN; 1.
DR PROSITE; PS00215; ADAM_MPRO; 1.
DR PROSITE; PS00427; Disintegrin_1; 1.
DR PROSITE; PS00214; Disintegrin_2; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Signal.
FT SIGNAL. 1 18 POTENTIAL.
FT CHAIN. 19 609 CROTICOLLASTATIN.
SQ SEQUENCE 609 AA; 68248 MW; D58876161F64FA5 CRC64;
Query Match 14.7%; Score 693.5; DB 13; Length 609;
Best Local Similarity 33.2%; Pred. No. 3.8e-43;
Matches 212; Conservative 106; Mismatches 276; Indels 45; Gaps 14;
QY 1 MLOGLPVSLLLSV---AVSAIKELPGVKYKVVVPIRLHLPHKREAKEPEQOQWETE 56
Db 1 MIQELL-VTICLAAPFYQSSIILESNGVNDYBIVPRKVTALPKGAV-----QPKYEDA 54
QY 57 LKYMTINGKIAVLKLNKKNLLAPGYTYTNSGTKEITTSPOIMWVYQGHILNEKV 116
Db 55 MQYELKNGEPVVLHKKNGKGFSEDIYSTHSPDQREITTPVEDHCHYHGRIONDAD 114
QY 117 SHASISTWRGLGYMSQGWRYWIEPLSPIHRWGQEHALWKNPWBKNYWSWTGMGVLW 176
Db 115 STASISACNGLAGHFKLQSEWYLIEPLKIPD--SEAHAVKYENVEKEDEALKMGVLT- 170
QY 177 AHWLQNTALPATKLVKLKRWKQOEHEKYLEYVLVWNGEWKRYNENQHEIRKRV 232
Db 171 QWESYEPIKASQV-----VTABHQKNPFRFVELFLVDKAMVTKNGDLKDKTKRM 225
QY 232 HEMANYVMYKKNLTHVALVGMETWTKWIKITPNASWTLNWSKWRGSLRRKRHW 291
Db 226 YEIVNTVNEIYRMYIHVALVGLIEWSNEDKLTVPFAGTTLNAPGWRKTDLLTKKH 285
QY 292 IAQLITATELAGTTVGLAWMTMSP--YSVGWQVMSNLLRVAGTMAHEMGNWGMW 350
Db 286 NQALLTAIDL-DRVIGLAVGSMCHPKRSTGIIQDYSEINLVVAVIMAHMGHNLGNE 344
QY 351 WYSWKWPSTIWMKALSWIPTWSSWSRLSYKWKWEKLSN---NLWNAPLPTWILT 407
Db 345 SGYCSGQYACIMRPEIS---PEPSTFFSNCSYFECWDFIMNHNPECIILNEPLGTDI 401
QY 408 PTWGNQVEMGEMWMTGTSEETNIWWAKTKIKATWQWALGEWKEKQWKKAGWVRP 467
Db 402 PFCGNELLEVBECDCGTPENCQECDAATCKLKSQSGCGHGDCCQCKFSKSGTECA 461
QY 468 AKWEMWLPDMWNGKSNWPMRWQVNGVPMWGHKGHLMTWPTLOEQWTELMGPTEVA 527
Db 462 SMSQCDPAHCTGQSSBCEPADVFHKNQGCPLDNYGYCNGNCPIMYHQCVDLFGADV 521
QY 528 WKSWYNRNGGSKYGYWRVWVWTLIPWKNWTKWGLWQGS---WNLPMKGRIVTWLTW 585
Db 522 EDSCFERNQKNYGYCRKNGNKIPCAPEDVKCGRLYCKDNPSPGQNNPCK----- 572
QY 586 KTWPEWTSQETGMVANGTKWKNKVNINAEVWLEKAY 624
Db 573 --MFYSNDEDEHKGWLPGLTKADGKVCNSNGHCHVDVATAY 609

RESULT 11
Q8JIR1 PRELIMINARY; PRT; 614 AA.
ID Q8JIR1
AC Q8JIR1;
DT 01-OCT-2002 (TremBLrel. 22, Created)
DT 01-OCT-2002 (TremBLrel. 22, Last sequence update)


```
Db 466 SECDAESCTGQADCTDDFKNGQPCLNHYGYCNGNCPIMEHQYALFGSNATVAED 525
QY 530 SWYNRRNGGSKYGYWRRVWMTLIPKANKWMTWQGLWQGGSWNLPPKGRIVTWLTWTKTW 589
Db 526 GCDEFNENGDKYFYCRKQSGVNPICAQEDVKQGLFCHCKKHPCDYK-----Y 573
QY 590 PENTSQIGVANGTKGWNKVINAEWVIEKAY 624
Db 574 SE--DPDYGVDNGTKCADGKVCNNGHCVDTATAY 606

RESULT 13
O93523 PRELIMINARY; PRT; 610 AA.
AC O93523; 1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Bothropasin (EC 3.4.24.49).
OS Bothrops jararaca (Jararaca).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Bothrops.
OX NCBI_TaxID=8724;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RA Asakura M.T., Camargo A.C.M., Serrano S.M.T.;
RT "Molecular cloning and sequence analysis of a cDNA encoding
RT bothropasin, a metalloproteinase isolated from the venom of Bothrops
RT jararaca."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF056025; AAC61986.2; -
DR HSP; P17494; 1KST.
DR MSRPS; M12.140; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR006586; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR001590; Peptidase M12B.
DR InterPro; IPR002870; Peptidase M12B_N.
DR Pfam; PF00200; disintegrin; 1.
DR Pfam; PF01562; Pep M12B propep; 1.
DR Pfam; PF01421; Reptolysin; 1.
DR PRINTS; PR00289; DISINTEGRIN.
DR ProDom; PD000664; Disintegrin; 1.
DR SMART; SM00608; ACR; 1.
DR SMART; SM00050; DISIN; 1.
DR PROSITE; PS0215; ADAM MEPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; 1.
DR PROSITE; PS0214; DISINTEGRIN_2; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase.
SQ SEQUENCE 610 AA; 68213 MW; 014C8AB6B6F25DD CRC64;

Query Match 14.5%; Score 684.5; DB 13; Length 610;
Best Local Similarity 32.1%; Pred No. 1.8e-42;
Matches 202; Conservative 112; Mismatches 278; Indels 37; Gaps 10;

QY 10 LLSVAVSA-----IKELPGVKYEVVYPIRLHPLHRAKSPQEQHETELKYON 61
Db 5 LLVTCIAAPFYQSGSIILSGNVNDYVYPRKVTPALPKGAV-----QPKYEDAMQVEF 59
QY 62 TINGKIADLYLKNKLLAPGTYEYNSGTKEITSPQIMWWWYQGHILNEKYSWASI 121
Db 60 KNGEPVVLHLEKNGKLFSDYSETHYSPDGSEITTPAVEDCHCYHGIENDADSTASI 119
QY 122 STWRGLRGVWSQGWRYTEPLSPHRCQEHALWKYNPWKNYSTWGMWGLWAFWLIQ 181
Db 120 SACNGLKGFKLQRETYFIEPLKLSN--SEAHAFKYNVEKDEAP-KXQGYT-QNWS 175
```

```
QY 182 QNTALPATKLVQLKWRKVQEBEKYIEYLYLVNGENKRYENQWEIRKRVEMANYVAML 241
Db 176 YEPIKASQSLVTTAEQOQYNPFYVVDQGMVTKQNGDLDKIKARMYELANIVNEI 235
QY 242 YKLTHTVALVGMIEITWTKWIKITPNASWTLXNWSKRGSVLSRRKHWHIAQLITATEL 301
Db 236 LRYLYMAALVGLIEINGDKITVKPDVDTYTLNSFAEWRKTDLLTRKKEHNRQLLTADF 295
QY 302 AGTTVGLAWSTWMSR-YSVGVVQWHSNLLRVAGTMAHENGHNWGMWHL---WYTSWKWP 357
Db 296 NGPTIGYAYIGSMCHPKRVAIVEDYSPINLVAVIMAHENGHNGLGIHHDIDFCSCGDYP 355
QY 358 STIWMKALSWYIPTWSSWSRLSYKWKWENKLSNWLWNAPLTWIILSTPLWGNQVEM 417
Db 356 C---IMGPTISNPSKFPFSCSYIQCDNFINKENPQCILNBEPLGTDIVSPVPCGNLELEV 412
QY 418 GEMWMTGTSBEWTNIWWAKTWIKATQWALGALGWEKWKQKAGWVRPAKWEWMLPEM 477
Db 413 GEECDGCTPENCQECDDAATCKLKSQCGHGDCECKFSKSGTECRASMECDPAEH 472
QY 478 WNGKSGNWPWRWQVNGWPHHGHKGMWLMGTWPTLQEQWTELMGPGTEVAVKSWYNRNEG 537
Db 473 CTGQSSECPADVTHKNGQPCLDNYGYCNGNCPIMYHQCVALFCADVYEAEDSCFKDNQK 532
QY 538 GSKYGYWRRVWMTLIPKANKWMTWQGLWQGG--WNLPWKGRIVTWLTWTKTWPEWTSQ 595
Db 533 GNYGYCRKENGKKIPCAPEDVKQGLYCKDNKSPGQNNPCK-----MFSYNDDE 581
QY 596 EIGWANGTKGWNKVINAEWVIEKAY 624
Db 582 HKGMVLPGTKADGKVCNNGHCVDTATAY 610

RESULT 14
Q8QG88 PRELIMINARY; PRT; 610 AA.
AC Q8QG88;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Metalloprotease.
OS Bothrops insularis (Island jararaca) (Queimada jararaca).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Bothrops.
OX NCBI_TaxID=8723;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RA Junqueira-de-Azevedo I.L.M., Ho P.L.;
RT "A survey of gene expression and diversity in the venom glands of the
RT pitviper Bothrops insularis through the generation of Expressed
RT Sequence Tags (ESTs).";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF490534; AAM09693.1; -
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR006586; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR001590; Peptidase M12B.
DR InterPro; IPR002870; Peptidase M12B_N.
DR Pfam; PF00200; disintegrin; 1.
DR Pfam; PF01562; Pep M12B propep; 1.
DR Pfam; PF01421; Reptolysin; 1.
DR PRINTS; PR00289; DISINTEGRIN.
DR ProDom; PD000664; Disintegrin; 1.
DR SMART; SM00608; ACR; 1.
DR PROSITE; PS0215; ADAM MEPRO; 1.
DR PROSITE; PS00427; DISINTEGRIN_1; 1.
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2004, 17:59:35 ; Search time 23 Seconds
(without alignments)
1739.571 Million cell updates/sec

Title: US-10-726-148A-15
Perfect score: 4712
Sequence: 1 MLOGLLPVSLLSVAISAIX.....PPTVKKWNPMTPKWSNEKA 775

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51525971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3128	66.4	775	4	US-09-786-256C-15
2	3128	66.4	775	4	US-09-786-256C-32
3	2132	45.2	540	4	US-09-786-256C-1
4	2132	45.2	540	4	US-09-786-256C-30
5	2079.5	44.1	529	2	US-08-836-442-3
6	817	17.3	201	4	US-09-786-256C-5
7	800.5	17.0	200	2	US-08-836-442-4
8	735.5	15.6	613	4	US-09-026-001A-10
9	732.5	15.5	621	4	US-09-026-001A-18
10	717	15.2	621	4	US-09-026-001A-6
11	702	14.9	470	1	US-08-514-014-2
12	702	14.9	470	2	US-08-833-823-2
13	690	14.6	470	3	US-08-813-150-2
14	690	14.6	470	4	US-09-546-553-2
15	674	14.3	592	4	US-09-026-001A-14
16	647	13.7	521	4	US-09-026-001A-12
17	635	13.6	616	4	US-09-608-790-1
18	596	12.6	439	4	US-09-026-001A-8
19	490.5	10.4	464	3	US-09-411-329C-14
20	490.5	10.4	464	4	US-09-846-729A-14
21	489.5	10.4	462	3	US-09-411-329C-3
22	489.5	10.4	462	3	US-09-411-329C-17
23	489.5	10.4	462	4	US-09-846-729A-3
24	489.5	10.4	462	4	US-09-846-729A-17
25	462.5	9.8	787	4	US-09-548-797B-5
26	462.5	9.8	802	4	US-09-632-098-2
27	462.5	9.8	812	4	US-09-632-098-4

28	462.5	9.8	849	4	US-09-548-797B-6
29	453	9.6	462	4	US-09-026-001A-16
30	443	9.4	746	4	US-09-548-797B-4
31	434	9.2	335	4	US-09-152-060-64
32	420.5	8.9	391	3	US-08-706-216-6
33	420.5	8.9	391	4	US-09-650-284B-6
34	405.5	8.6	855	3	US-09-813-819-2
35	405.5	8.6	855	4	US-09-920-048-2
36	401.5	8.5	814	3	US-09-813-819-4
37	401.5	8.5	814	4	US-09-920-048-4
38	397.5	8.4	751	2	US-08-836-443-3
39	334.5	7.1	240	4	US-09-152-060-100
40	326.5	6.9	781	4	US-09-738-946-8
41	290	6.2	96	4	US-09-786-256C-2
42	284	6.0	215	2	US-08-836-443-4
43	265.5	5.6	201	3	US-09-411-329C-1
44	265.5	5.6	201	4	US-09-411-335-1
45	265.5	5.6	201	4	US-09-466-276-1

ALIGNMENTS

RESULT 1
US-09-786-256C-15
; Sequence 15, Application US/09786256C
; Patent No. 6680189
; GENERAL INFORMATION:
; APPLICANT: YOSHIMURA, Koji
; APPLICANT: HIKICHI, Yuichi
; APPLICANT: NISHIMURA, Atsushi
; TITLE OF INVENTION: No. 6680189el Protein and DNA Thereof
; FILE REFERENCE: 2544 USOP
; CURRENT APPLICATION NUMBER: US/03/786,256C
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: PCT/JP99/04766
; PRIOR FILING DATE: 1999-09-02
; PRIOR APPLICATION NUMBER: JP 10-250115
; PRIOR FILING DATE: 1998-09-03
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 775
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(775)
; OTHER INFORMATION: An isolated ADAM family protein
US-09-786-256C-15

Query Match	66.4%	Score 3128;	DB 4;	Length 775;
Best Local Similarity	84.4%	Pred. No. 5.7e-268;		
Matches	654;	Conservative 29;	Mismatches 92;	Indels 0; Gaps 0;
Qy	1	MLOGLLPVSLLSVAISAIXELPGVKYVYVPIRLHPLHKRAKEPEQOQWETELKYK	60	
Db	1	MLOGLLPVSLLSVAISAIXELPGVKYVYVPIRLHPLHKRAKEPEQOQWETELKYK	60	
Qy	61	MTINGKIAVLYLKQKNLLAPGYTETYNSTGKEITTSFQIMDDCYQGHILKEKYSWAS	120	
Db	61	MTINGKIAVLYLKQKNLLAPGYTETYNSTGKEITTSFQIMDDCYQGHILKEKYSWAS	120	
Qy	121	ISTWRGLRYNSQHQRYWIEPLSPHWRQOEHALWKYPKWKYNSYTWGNCVLAHNL	180	
Db	121	ISTWRGLRYNSQHQRYWIEPLSPHWRQOEHALWKYPKWKYNSYTWGNCVLAHNL	180	
Qy	181	QONIALPATKVLKWKVKQEHKEKYEYVYLWNGEMKRYNENQWIRKRVWEMANYNM	240	
Db	181	QONIALPATKVLKWKVKQEHKEKYEYVYLWNGEMKRYNENQWIRKRVWEMANYNM	240	
Qy	241	LYKKLNTHVALVQMEITWTKWIKITPPNASWTLENMSKWRGVSLSRRKRWIAOLITATE	300	
Db	241	LYKKLNTHVALVQMEITWTKWIKITPPNASWTLENMSKWRGVSLSRRKRWIAOLITATE	300	

OTHER INFORMATION: An isolated ADAM family protein
US-09-786-256C-1

```
Query Match      45.2%; Score 2132; DB 4; Length 540;
Best Local Similarity 84.9%; Pred. No. 3.8e-180;
Matches 444; Conservative 20; Mismatches 59; Indels 0; Gaps 0;

QY 1 MLOGLLPVSLLSVAVAISAKELPGVKYEVVYPIRLHPLHKREAKEPEQOQOWETELKYK 60
DB 1 MLOGLLPVSLLSVAVAISAKELPGVKYEVVYPIRLHPLHKREAKEPEQOQOWETELKYK 60
QY 61 MTINKIAVLKXKNKLLAPGYTTYNTSTGKEITTSPTQIMDDCYOQHILNEKVSAS 120
DB 61 MTINKIAVLKXKNKLLAPGYTTYNTSTGKEITTSPTQIMDDCYOQHILNEKVSAS 120
QY 121 ISTWGLAGYVSQGWRYWIEPLSPIHRWGQSHALWKYNPWEKKNYKWTMGWGLVAHNL 180
DB 121 ISTWGLAGYVSQGWRYWIEPLSPIHRWGQSHALWKYNPWEKKNYKWTMGWGLVAHNL 180
QY 181 QONIALPATKLVKLRKVKQVQEHKEIEYLVLMNGEWKRYNENQWEIRKRVWEMANYNM 240
DB 181 QONIALPATKLVKLRKVKQVQEHKEIEYLVLMNGEWKRYNENQWEIRKRVWEMANYNM 240
QY 241 LYKLNTHVALVGMIEIWTDKIKITPNSFTLENNSKWRGSLRRKHHTIAQLITATE 300
DB 241 LYKLNTHVALVGMIEIWTDKIKITPNSFTLENNSKWRGSLRRKHHTIAQLITATE 300
QY 301 LAGTTVGLAWSTWSPYSVGVVQVHSDNLLRVAGTMAHEMGNHMMHWYSKWPSTI 360
DB 301 LAGTTVGLAWSTWSPYSVGVVQVHSDNLLRVAGTMAHEMGNHMMHWYSKWPSTI 360
QY 361 WMMKALSWYIPTWSSWSRLSYKWKWMLNNAELPTWIIISTPIGNQVEMGEW 420
DB 361 WMMKALSWYIPTWSSWSRLSYKWKWMLNNAELPTWIIISTPIGNQVEMGEW 420
QY 421 WWWGTSEBNTNIIWAKTKIKATQWALGEWKKWKKAGVYKVRPAKWEWLPWMNG 480
DB 421 WWWGTSEBNTNIIWAKTKIKATQWALGEWKKWKKAGVYKVRPAKWEWLPWMNG 480
QY 481 KSGNPPWRWQVNGWPHHGHGHLMTPTLQEQWTELWGP 523
DB 481 KSGNPPWRWQVNGWPHHGHGHLMTPTLQEQWTELWGP 523
```

```
RESULT 4
US-09-786-256C-30
; Sequence 3C, Application US/09786256C
; Patent No. 6680189
; GENERAL INFORMATION:
; APPLICANT: YOSHIMURA, Koji
; APPLICANT: HIKICHI, Yoichi
; APPLICANT: NISHIMURA, Atsushi
; TITLE OF INVENTION: No. 6680189el Protein: and DNA Thereof
; FILE REFERENCE: 2544 USOP
; CURRENT APPLICATION NUMBER: US/09/786,256C
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: PCT/JP99/04766
; PRIOR FILING DATE: 1999-09-02
; PRIOR APPLICATION NUMBER: JP 10-250115
; PRIOR FILING DATE: 1998-09-03
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 11..(2560)
; OTHER INFORMATION: DNA sequence of FIG 1-2 containing SEQ ID NO:3 encoding for prote
; OTHER INFORMATION: in of SEQ ID NO: 1
US-09-786-256C-30
```

```
Query Match      45.2%; Score 2132; DB 4; Length 540;
Best Local Similarity 84.9%; Pred. No. 3.8e-180;
Matches 444; Conservative 20; Mismatches 59; Indels 0; Gaps 0;

QY 1 MLOGLLPVSLLSVAVAISAKELPGVKYEVVYPIRLHPLHKREAKEPEQOQOWETELKYK 60
DB 1 MLOGLLPVSLLSVAVAISAKELPGVKYEVVYPIRLHPLHKREAKEPEQOQOWETELKYK 60
QY 61 MTINKIAVLKXKNKLLAPGYTTYNTSTGKEITTSPTQIMDDCYOQHILNEKVSAS 120
DB 61 MTINKIAVLKXKNKLLAPGYTTYNTSTGKEITTSPTQIMDDCYOQHILNEKVSAS 120
QY 121 ISTWGLAGYVSQGWRYWIEPLSPIHRWGQSHALWKYNPWEKKNYKWTMGWGLVAHNL 180
DB 121 ISTWGLAGYVSQGWRYWIEPLSPIHRWGQSHALWKYNPWEKKNYKWTMGWGLVAHNL 180
QY 181 QONIALPATKLVKLRKVKQVQEHKEIEYLVLMNGEWKRYNENQWEIRKRVWEMANYNM 240
DB 181 QONIALPATKLVKLRKVKQVQEHKEIEYLVLMNGEWKRYNENQWEIRKRVWEMANYNM 240
QY 241 LYKLNTHVALVGMIEIWTDKIKITPNSFTLENNSKWRGSLRRKHHTIAQLITATE 300
DB 241 LYKLNTHVALVGMIEIWTDKIKITPNSFTLENNSKWRGSLRRKHHTIAQLITATE 300
QY 301 LAGTTVGLAWSTWSPYSVGVVQVHSDNLLRVAGTMAHEMGNHMMHWYSKWPSTI 360
DB 301 LAGTTVGLAWSTWSPYSVGVVQVHSDNLLRVAGTMAHEMGNHMMHWYSKWPSTI 360
QY 361 WMMKALSWYIPTWSSWSRLSYKWKWMLNNAELPTWIIISTPIGNQVEMGEW 420
DB 361 WMMKALSWYIPTWSSWSRLSYKWKWMLNNAELPTWIIISTPIGNQVEMGEW 420
QY 421 WWWGTSEBNTNIIWAKTKIKATQWALGEWKKWKKAGVYKVRPAKWEWLPWMNG 480
DB 421 WWWGTSEBNTNIIWAKTKIKATQWALGEWKKWKKAGVYKVRPAKWEWLPWMNG 480
QY 481 KSGNPPWRWQVNGWPHHGHGHLMTPTLQEQWTELWGP 523
DB 481 KSGNPPWRWQVNGWPHHGHGHLMTPTLQEQWTELWGP 523
```

```
RESULT 5
US-08-836-442-3
; Sequence 3, Application US/08836442
; Patent No. 5990293
; GENERAL INFORMATION:
; APPLICANT: DOCHERTY, Andrew, J.P.
; APPLICANT: SLOCOMBE, Patrick, M.
; TITLE OF INVENTION: A HUMAN METALLOPROTEINASE
; TITLE OF INVENTION: VARIANTS THEREOF AND DNA SEQUENCES CODING THEREFOR
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,442
; FILING DATE: 01-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB96/02181
; FILING DATE: 13-MAR-1997
; APPLICATION NUMBER: GB 9612150.4
; FILING DATE: 11-JUN-1996
; APPLICATION NUMBER: GB 9526229.1
```

```

; FILING DATE: 21-DEC-1995
; APPLICATION NUMBER: GB 9521498.7
; FILING DATE: 23-OCT-1995
; APPLICATION NUMBER: GB 95521495.3
; FILING DATE: 20-OCT-1995
; APPLICATION NUMBER: GB 9518023.8
; FILING DATE: 05-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Resnick, David S
; REGISTRATION NUMBER: 34,235
; REFERENCE/DOCKET NUMBER: 47425
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5990293e
US-08-836-442-3

Query Match 44.1%; Score 2079.5; D3 2; Length 529;
Best Local Similarity 83.6%; Pred. No. 1.6e-175;
Matches 437; Conservative 20; Mismatches 55; Indels 11; Gaps 2;

QY 1 MLOGLPVSLLSVSAISKELPGVKYEVVYPIRLHPLHKAKEPEQEOQWETELKYK 60
DB 1 MLOGLPVSLLSVSAISKELPGVKYEVVYPIRLHPLHKAKEPEQEQFETELKYK 60

QY 61 MTINGKIAVLKKNKXLLAPGYETBYNSTGKEITTSPOIMWYVYQGHILNEKYSWAS 120
DB 61 MTINGKIAVLKKNKXLLAPGYETBYNSTGKEITTSPOIMDCYQGHILNEKYSWAS 120

QY 121 ISTWRGLRGVWSQGWRYETPLSPHRCQEHALWKYPWEKNYKSTGCMGWVLAHNL 180
DB 121 ISTCRGLRGVWSQGWRYETPLSPHRCQEHALWKYPWEKNYKSTGCMGWVLAHNL 180

QY 181 QONIALPATKLVKLKWKVQEHKYEYLVLMNGEKRYNENQWEIRKRVWEMANYNM 240
DB 181 QONIALPATKLVKLKWKVQEHKYEYLVLMNGEKRYNENQWEIRKRVWEMANYNM 240

QY 241 LYKLNTHVALVGMETWTKWTKITPNASWTLENWSKWRGSLSRKRHWIAQLITATE 300
DB 241 LYKLNTHVALVGMETWTKWTKITPNASWTLENWSKWRGSLSRKRHWIAQLITATE 300

QY 301 LAGTTVGLAWKMTWMSYSGVVOHWSNLLRVAGTMAHEMGNHGMWMMWYKWKWPSTI 360
DB 301 LAGTTVGLAWKMTWMSYSGVVOHWSNLLRVAGTMAHEMGNHGMWMMWYKWKWPSTI 360

QY 361 WNMKALSWIPTWMSWSLSYKWKWEMKLSNMLNAPLPTWIISTPIWGNQLVEMGEW 420
DB 361 WNMKALSWIPTWMSWSLSYKWKWEMKLSNMLNAPLPTWIISTPIWGNQLVEMGEW 420

QY 421 WNGTSEWNTNWNNAKTKWKATQWALGEWKEKQWKGKAGWVRPAKWEMWLPENWNG 480
DB 421 WNGTSEWNTNWNNAKTKWKATQWALGEWKEKQWKGKAGWVRPAKWEMWLPENWNG 480

QY 481 XSGNPNWNRQVGNWPNHKGHWLMTGPTLQBOQWTELWCPG 523
DB 471 XSGNCPDPRFQVNGFPC-HKGHGLMGTCTPLQBOQWTELWCPG 512

RESULT 6
US-09-786-256C-5
; Sequence 5, Application US/09786256C
; Patent No. 6680189
; GENERAL INFORMATION:
; APPLICANT: YOSHIMURA, Koji
; APPLICANT: HIKICHI, Yuichi
; APPLICANT: NISHIMURA, Acsushi

; TITLE OF INVENTION: No. 6680189el Protein and DNA Thereof
; FILE REFERENCE: 2544 USOP
; CURRENT APPLICATION NUMBER: US/03/786,256C
; CURRENT FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: PCT/JP99/04766
; PRIOR FILING DATE: 1999-09-02
; PRIOR APPLICATION NUMBER: JP 10-250115
; PRIOR FILING DATE: 1998-09-03
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 5
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(201)
; OTHER INFORMATION: Isolated fragment of protein of SEQ ID NO.1 (aa 199-399)
US-09-786-256C-5

Query Match 17.3%; Score 817; DB 4; Length 201;
Best Local Similarity 84.6%; Pred. No. 1.7e-64;
Matches 170; Conservative 12; Mismatches 19; Indels 0; Gaps 0;

QY 199 VQSEHKYIEYLVLMNGEKRYNENQWEIRKRVWEMANYNMVLYKLNTHVALVGMETW 258
DB 1 VQSEHKYIEYLVLMNGEKRYNENQWEIRKRVWEMANYNMVLYKLNTHVALVGMETW 258

QY 259 WKWKIITPNASWTLENWSKWRGSLSRKRHWIAQLITATELAGTTVGLAWKMTWMSY 318
DB 61 DKDKIITPNASWTLENWSKWRGSLSRKRHWIAQLITATELAGTTVGLAWKMTWMSY 318

QY 319 SVGVVOHWSNLLRVAGTMAHEMGNHGMWMMWYKWKWPSTIWMKALSWIPTWMSWS 378
DB 121 SVGVVOHWSNLLRVAGTMAHEMGNHGMWMMWYKWKWPSTIWMKALSWIPTWMSWS 378

QY 379 SRLSYKWKWEMKLSNMLNAP 399
DB 181 SRLSYKWKWEMKLSNMLNAP 201

RESULT 7
US-08-836-442-4
; Sequence 4, Application US/08836442
; Patent No. 5990293
; GENERAL INFORMATION:
; APPLICANT: DOCHERTY, Andrew, J.P.
; APPLICANT: SLOCOMBE, Patrick, M.
; TITLE OF INVENTION: A HUMAN METALLOPROTEINASE
; TITLE OF INVENTION: VARIANTS THEREOF AND DNA SEQUENCES CODING THEREFOR
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,442
; FILING DATE: 01-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB96/02181
; FILING DATE: 13-MAR-1997
; APPLICATION NUMBER: GB 9612150.4
; FILING DATE: 11-JUN-1996
; APPLICATION NUMBER: GB 9526229.1

```

```
;
; FILING DATE: 21-DEC-1995
; APPLICATION NUMBER: GB 9521498.7
; FILING DATE: 20-OCT-1995
; APPLICATION NUMBER: GB 9521495.3
; FILING DATE: 20-OCT-1995
; APPLICATION NUMBER: GB 9518023.8
; FILING DATE: 05-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Resnick, David S
; REGISTRATION NUMBER: 34,235
; REFERENCE/DOCKET NUMBER: 47425
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 200 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5990293e
; US-08-836-442-4
;
; Query Match 17.0%; Score 800.5; DB 2; Length 200;
; Best Local Similarity 84.1%; Pred. No. 4.8e-63;
; Matches 169; Conservative 12; Mismatches 13; Indels 1; Gaps 1;
;
QY 199 VQEHKYEYLVLMNGEWEKRYNENQWEIRKRYMANYVNMVYKKNLTHVALVGMETWT 258
DB 1 VQEHKYEYLVLDN-EFKRYNENQDEIRKRYEMANYVNMVYKKNLTHVALVGMETWT 59
;
QY 259 WKWKIKTPNASWTLNENKRGVSVLSRRKHWHIAQLTATLACTTGVGLAMWMTWSPY 318
DB 60 DKDKIKTPNASFTLENFSKWRGVSLSRRKHWHIAQLTATLACTTGVGLAMWMTWSPY 119
;
QY 319 SVGVVQHSNLLRVAGTMAHEMGNKSWWYKSWKWPSTIWMVKALSWYIPTWSSW 378
DB 120 SVGVVQHSNLLRVAGTMAHEMGNHFNFGHDDYCKPCSTICVMDKALSFYIPTDPS 179
;
QY 379 SRLSYKWKWENKLSNWLWNP 399
DB 180 SELSYDKFFEDKLSNCLFNP 200
;
RESULT 8
US-09-026-001A-10
; Sequence 10, Application US/09026001A
; Patent No. 6413760
; GENERAL INFORMATION:
; APPLICANT: Boonhoo, Amechand
; APPLICANT: Seehra, Jasbir
; APPLICANT: Shaw, Gray
; APPLICANT: Sako, Dianne
; TITLE OF INVENTION: HIGHLY PURIFIED MOCAHAGIN, A COBRA VENOM
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/026,001A
; FILING DATE: 18-FEB-1998
; CLASSIFICATION: 435
;
; FILING DATE: 21-DEC-1395
; APPLICATION NUMBER: GB 9521498.7
; FILING DATE: 20-OCT-1995
; APPLICATION NUMBER: GB 9521495.3
; FILING DATE: 20-OCT-1995
; APPLICATION NUMBER: GB 9518023.8
; FILING DATE: 05-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A. 32,724
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15293B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 613 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-026-001A-10
;
; Query Match 15.6%; Score 735.5; DB 4; Length 613;
; Best Local Similarity 33.8%; Pred. No. 1.2e-56;
; Matches 214; Conservative 109; Mismatches 280; Indels 31; Gaps 10;
;
QY 1 MLOGLLPVSLLSV---AVSAIKELPGYKVVVYPIRLHPLHREAKEPEQOEOWETE 56
DB 1 MIQALL-VAICLAVFPYOGSSIIILSGNVNDYEVVYPOKPSLKGVGVOPOPEYKIE 59
;
QY 57 LKYMNTINGKIAVLYLKKNNKLLAPGYTETYNSTGKEITTSFQIMMMWYIYQGHILNEKV 116
DB 60 MYPEQVNGEPVVLHLERNKGLFSEDTETHYASDGREITTSPLVQDHCYHYGIONEAD 119
;
QY 117 SWASTSTWRGLRGYSQGWRYMIEPLSPIHMGQEHALMKYNPWEKONTWSTGKMGVJLM 176
DB 120 SSAVISACDGLKGHELOQETFYIPLKISD--SEAHAIYKDNVE-NEDETPTCGVTE 176
;
QY 177 AHWLQONLALPATKLVKL-KWRKVOHEKYIEYLVLMNGEWEKRYNENQWEIRKRYEMA 235
DB 177 TTWSEDESIEKTSQNTPTPEQRYIQAOKKYLEFYVVVDNMTYKRYKRPVILKRRVIE 236
;
QY 236 NYVNLKYNLTHVALVGMETWTWKIKITPNASWTLNENKRGVSVLSRRKHWHIAQL 295
DB 237 NPMNMYNLFHIALIGLEIWSNENEINVQSDVQATLDLFGEWREKLLPKRNDNAQL 296
;
QY 296 IPATELAGTTVGLAMWMTWSP-YGVVQVQWISNLLRVAGTMAHEMGNKSWWYKSW 354
DB 297 LTGIDFKGTPVGLAYIGSICNPKSSVAVVQDYSSRTSMVAITMAHEMGNMGIIHEDGPSC 356
;
QY 355 KWPSTIWMVKALSWYIPTM-WSSASRLSYKWKWENKLSNWLWNPALPTWISTPIWGNQ 413
DB 357 TCGSNKCVN--STRTEPAYQFSVCSVREHQBILRRDPCQILNKLPSTDIIVSPPICGN 414
;
QY 414 LVEMGEWMMWGTSEBWTIWMWAKTWIKATQWALGEWKEWQWQKAGWVRPAKWEMW 473
DB 415 FVEVGECDGSPADCSACCDATTCKLQPHAQDSEGCCEKCKPKGAGACRAAKDCD 474
;
QY 474 LPEMNGKSGNPMWRQVNGWPHHKGHLMGWTPTLQBOQTEWLGSTEVAWKSWYN 533
DB 475 LPELCTGQSAECFTDIFORGLPCQNNEGCYNGKCPIMTNQCIALLRGVGVKVSRCFT 534
;
QY 534 RNEGSKYGYRWRVWMTLIPKANMTWGLKAWQGS---WNLPWKGRIVTWLTWIKWNP 593
DB 535 LMQRTSGGGLCRMEYGRKIPCAAKDVCKGRLLKKGNSMICNCSVSPR-----592
;
QY 591 EWTQSIGMVGNTKGMKNKVINAEWVMEKAY 624
DB 583 ---DPSYGNVPEGTGCGDMVCSNQCVDVKTAY 613
;
RESULT 9
US-09-026-001A-18
; Sequence 18, Application US/09026001A
; Patent No. 6413760
; GENERAL INFORMATION:
; APPLICANT: Boonhoo, Amechand
; APPLICANT: Seehra, Jasbir
; APPLICANT: Shaw, Gray
;
; FILING DATE: 21-DEC-1395
; APPLICATION NUMBER: GB 9521498.7
; FILING DATE: 20-OCT-1995
; APPLICATION NUMBER: GB 9521495.3
; FILING DATE: 20-OCT-1995
; APPLICATION NUMBER: GB 9518023.8
; FILING DATE: 05-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Resnick, David S
; REGISTRATION NUMBER: 34,235
; REFERENCE/DOCKET NUMBER: 47425
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 200 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5990293e
; US-08-836-442-4
;
; Query Match 17.0%; Score 800.5; DB 2; Length 200;
; Best Local Similarity 84.1%; Pred. No. 4.8e-63;
; Matches 169; Conservative 12; Mismatches 13; Indels 1; Gaps 1;
;
QY 199 VQEHKYEYLVLMNGEWEKRYNENQWEIRKRYMANYVNMVYKKNLTHVALVGMETWT 258
DB 1 VQEHKYEYLVLDN-EFKRYNENQDEIRKRYEMANYVNMVYKKNLTHVALVGMETWT 59
;
QY 259 WKWKIKTPNASWTLNENKRGVSVLSRRKHWHIAQLTATLACTTGVGLAMWMTWSPY 318
DB 60 DKDKIKTPNASFTLENFSKWRGVSLSRRKHWHIAQLTATLACTTGVGLAMWMTWSPY 119
;
QY 319 SVGVVQHSNLLRVAGTMAHEMGNKSWWYKSWKWPSTIWMVKALSWYIPTWSSW 378
DB 120 SVGVVQHSNLLRVAGTMAHEMGNHFNFGHDDYCKPCSTICVMDKALSFYIPTDPS 179
;
QY 379 SRLSYKWKWENKLSNWLWNP 399
DB 180 SELSYDKFFEDKLSNCLFNP 200
;
RESULT 8
US-09-026-001A-10
; Sequence 10, Application US/09026001A
; Patent No. 6413760
; GENERAL INFORMATION:
; APPLICANT: Boonhoo, Amechand
; APPLICANT: Seehra, Jasbir
; APPLICANT: Shaw, Gray
; APPLICANT: Sako, Dianne
; TITLE OF INVENTION: HIGHLY PURIFIED MOCAHAGIN, A COBRA VENOM
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/026,001A
; FILING DATE: 18-FEB-1998
; CLASSIFICATION: 435
;
```

APPLICANT: Sako, Dianne
TITLE OF INVENTION: HIGHLY PURIFIED MOCARHAGIN, A COBRA VENOM
TITLE OF INVENTION: THERAPEUTIC USES THEREOF
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/026,001A
FILING DATE: 18-FEB-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15293B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 621 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-026-001A-18

Query Match 15.5%; Score 732.5; DB 4; Length 621;
Best Local Similarity 33.5%; Pred. No. 2.2e-56;
Matches 216; Conservative 108; Mismatches 277; Indels 43; Gaps 11;
Qy 1 MLOGLPVSLLSV-----AVSAIKELPGVKYEVVYPIRLHPLHKREAKEPEQSQWETE 56
Db 1 MIOALL-VAICLAVFPYQSSIIIESGNVDYEVVYPOKVPALSKGGVQNEQPTKYBDT 59
Qy 57 LKYKMTNGKAVLVLKKNKLLAPGYTETYNSTGKEITTSPOIMWWWYQGHILNEKV 116
Db 60 MQYEFQNGEPVVLHLRKNKGLFSDYTHYAPGGRITTSPPVDHCYHYGIONEAD 119
Qy 117 SWASISTWRGLRGYSQGWRYWIEPLSPIHRWQGEHALWKYNPWKNYWTGMMGVLM 176
Db 120 SSAVISACDGLKGFELQGETYFIEPLKISD--SEAHAIYKDNVE-NEDETPETCGYTE 176
Qy 177 AHW-----LQONIALPATKVLKWKVKQEHKEKYEYLVLVNKGEMKRYENQW 225
Db 177 TTWESDESIEKTSQDDDKRPPTNPQD--RYLQAKKYLEFVVVVMINIRHYKDKP 234
Qy 226 EIRKRVEMANVNLKLLMTHVALGMEIWTWKIKITENASWTLENKSKRGSVLS 285
Db 235 VIKRVRVEMINMNVNRLFHIALIGLEINSRNEINVOQSDVQATLDLFGWEKXLL 294
Qy 286 RKRHWIAQLITATLACTTGLAWSTMWGP-YSGVGVQVHSMNLLRVAGTMAHEMGN 344
Db 295 PRKRNDAQLLTGIDFKGTPGLAVIGSICNPKSSVAVVQDYSRTSMVAITMAHEMGN 354
Qy 345 KGMWHWYTSKWPSIWMKALSWYITW-MSSWSRLSYKWKWENKLSNMLNAPLPTW 403
Db 355 MGIHDSGSCSGSKKCYM--STRTEPAYQFSSCSVREHBYLJDRPQCILNPLSTD 412
Qy 464 IISTPIWGNQVEMGEWMMGTSEBWTNLWNAKWKIKATQWALGEMWKEWKKQKQM 463
Db 413 VSPPIGNNFNVEGEECDGSPADQSCACCDATTCKLPHQACDSGECCECKFKGAGA 472
Qy 464 VWRPAKWEWLPENWNGKSGNPNRWNQVNGPWHGKHLMGTWPTLQEQMTLWGP 523

Db 473 ECRAAKDDCDLPELCTGSAECPTDIFQNGLPQONNEGKYNGKCFIMTNQCIALRPG 532
Qy 524 TEVAWSWYNRNEGSKYGVRRVWVTLIPKAWMTMWKGLWQGS---WNLPWKGRIV 580
Db 533 KVVRSDSCFTLNQRTSGGLCRMEYGRKIPCAAKDVYKGRLLFCCKGSMICNCSVSR-- 590
Qy 581 TWLTWKTWPMTSQSIGMVANGTKWGNKVKWNAEAWVWIEKAY 624
Db 591 -----DPSYGMVPEPTKCGDGMVCSNRQCVDVKTAY 621
RESULT 10
US-09-026-001A-6
Sequence 6, Application US/09026001A
Patent No. 6413760
GENERAL INFORMATION:
APPLICANT: Boodhoo, Anechand
APPLICANT: Seehra, Jasbir
APPLICANT: Shaw, Gray
APPLICANT: Sako, Dianne
TITLE OF INVENTION: HIGHLY PURIFIED MOCARHAGIN, A COBRA VENOM
TITLE OF INVENTION: THERAPEUTIC USES THEREOF
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/026,001A
FILING DATE: 18-FEB-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15293B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 621 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-026-001A-6

Query Match 15.2%; Score 717; DB 4; Length 621;
Best Local Similarity 35.4%; Pred. No. 5.1e-55;
Matches 196; Conservative 98; Mismatches 232; Indels 28; Gaps 11;
Qy 1 MLOGLPVSLLSV-----AVSAIKELPGVKYEVVYPIRLHPLHKREAKEPEQSQWETE 56
Db 1 MIOALL-VAICLAVFPYQSSIIIESGNVDYEVVYPOKVPALSKGGVQNPQETKYEDT 59
Qy 57 LKYKMTNGKAVLVLKKNKLLAPGYTETYNSTGKEITTSPOIMWWWYQGHILNEKV 116
Db 60 MQYEFQNGEPVVLHLRKNKGLFSDYTHYAPGGRITTSPPVDHCYHYGIONEAD 119
Qy 117 SWASISTWRGLRGYSQGWRYWIEPLSPIHRWQGEHALWKYNPWKNYWTGMMGVLM 176
Db 120 SSAVISACDGLKGFELQGETYFIEPLKISD--SEAHAIYKDNVEEER-EIPKICGVTO 176
Qy 177 AHWLQONIALPATKVLKWKVKQEH-----EKYIEYLVLVNKGEMKRYENQWIRKRYW 232

Db 177 TTWESDE---PIEKSSQLTNTPEQRYLOAKYIBFYVVVDNWTFRKXTGKHJVTTRVY 233
 QY 233 EMANYVNMVLYKLNTHVALVGMIEIWTWKIKITPNASWTLNEMSKWGSVLSRRKRHWI 292
 Db 234 EYVNAINTYXRELPHIALIGLEIWSNGEINVOQSDVQATLDFGEWRENKLLPRKNDN 293
 QY 293 AQLTATLATEAGTTCGLAMWMTWSP-YSGVGVQVHSHNLLRVAGTMAHEMGNHGMHW 351
 Db 294 AQLTATSTEFNGTTTGLGIGISLSPKSVAVYQDHSKSTSMVAITVAHQMGHNLGNDDR 353
 QY 352 YSKWKPSTIWMWKAISVITPWSWSRLSFKWKEKLSN---WLNAPLPTWLIISTP 408
 Db 354 ASCTCGSKNCLN--STKYIESL--SEFSSCSQVEHREYLLRDRPOCILNPKRKAIVTPP 409
 QY 409 IWGNOLVEMGEWMMWGTSTSEEMNIWMAKTWKIKATQWALGEWKEWKEWKEWKEW 468
 Db 410 VCGNYFVERGEBECGSPEDCQNTCCDAATCKLQHEAQCDSECECKCKFKGAGAECAA 469
 QY 469 KHEWLPENKNGSKNMPWVRQVNGWPHHKGHGLWGTWTLQBOITELMGPGTEVA- 527
 Db 470 KNDCCFPELCTGRSAKPRDSFQRNGHPQNNQGYCYNGTCPTLTNQCATLWGPCKMSP 529
 QY 528 ---WKSNNRNEG 537
 Db 530 GLCFMKNWNRSCG 543

RESULT 11

US-08-514-014-2
 ; Sequence 2, Application US/08514014
 ; Patent No. 5707829
 ; GENERAL INFORMATION:
 ; APPLICANT: Jacobs, Kenneth
 ; APPLICANT: McCoy, John
 ; APPLICANT: Kelleher, Kerry
 ; APPLICANT: Carlin, McKeough
 ; TITLE OF INVENTION: DNA SEQUENCES AND SECRETED PROTEINS
 ; TITLE OF INVENTION: ENCODED THEREBY
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSER: Genetics Institute, Inc. -- Legal Affairs
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/514,014
 ; FILING DATE:
 ; CLASSIFICATION: 514
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brown, Scott A.
 ; REGISTRATION NUMBER: 32,724
 ; REFERENCE/DOCKET NUMBER: GI6000
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 498-8224
 ; TELEFAX: (617) 876-5851
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 470 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-514-014-2

Query Match 14.9%; Score 702; DB 1; Length 470;
 Best Local Similarity 39.2%; Pred. No. 7.5e-54;

Matches 177; Conservative 87; Mismatches 161; Indels 26; Gaps 8;
 QY 5 LLPVS-LLLSVAVSAIKELPGVKVEVYVPIRLHPLKKEAPE-----QQQWTELK 58
 Db 18 LLPVLWLIYVQTAIAIKQTPELTHBIIVCPKHLHLKKEBIKNNOTKHKGEERYPEVQ 77
 QY 59 YKQTINGKIAVLVYKKNKOLLAPGYTETVYNGSTGKEITTSPOIMWVYVQGHILNEKVS 118
 Db 78 YQMLNGEBEILLSLQTKHLLQPDYTETLYSPRGEIITTKPENMEHCYTKENILNEKNSV 137
 QY 119 ASISWIRGLRGYSQGWQWYIEPLSPIHRWQOEHALMKYNPWEKN-----YWSWGM 171
 Db 138 ASISCDGLRGYPTTHHQRYQIKPLKSTDE--KEHAVFTSNQOEQDPANHTCGVKSTGK 195
 QY 172 WGVLAHMTQQNIALPATKLVKWKVQVEHKEYIEYLVLVNNGEKKRYNEQWETRKRV 231
 Db 196 QGPI---RISRLKSP-----KEDELRAQ---KYIDLVLVDNAFYKNYNNENLTLSRFV 245
 QY 232 WEMANVNMVLYKLNTHVALVGMIEIWTWKIKITPNASWTLNEMSKWGSVLSRRKRHW 291
 Db 246 FVMMNLNVYNTIDVQVALVGMIEINSDKIKVPSASTTFDNFLRWSSNLG-KKIID 304
 QY 292 IAQLITATELAGTTCGLAMWMTWSPYSVGVQVHSHNLLRVAGTMAHEMGNHGMHW 351
 Db 305 HAQLLSGISFNRRVGLAASNSLSPSSVAVIEAKKNVVALVGVMSHELGHVGLGMPDVP 364
 QY 352 YSKWKPSTIWMWKAISVITPWSWSRLSFKWKEKLSNWLWNAPLPTWLIISTPIMG 411
 Db 365 FNTKCPSGCVNMVYLSKPKPDKFSTSCRAHFERVLLSOKPKCLLQAPITNIMTTPVCG 424
 QY 412 NOLVEMGEWMMWGTSTSEEMNIWMAKTWKIK 442
 Db 425 NHLLEVGEBCDCGSPKCTNLCCEALTCKLK 455

RESULT 12

US-08-833-823-2
 ; Sequence 2, Application US/08833823
 ; Patent No. 5969093
 ; GENERAL INFORMATION:
 ; APPLICANT: Jacobs, Kenneth
 ; APPLICANT: McCoy, John
 ; APPLICANT: Kelleher, Kerry
 ; APPLICANT: Carlin, McKeough
 ; TITLE OF INVENTION: DNA SEQUENCES AND SECRETED PROTEINS
 ; TITLE OF INVENTION: ENCODED THEREBY
 ; NUMBER OF SEQUENCES: 12
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSER: Genetics Institute, Inc. -- Legal Affairs
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/833,823
 ; FILING DATE: 10-APR-1997
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/514,014
 ; FILING DATE: 11-AUG-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brown, Scott A.
 ; REGISTRATION NUMBER: 32,724
 ; REFERENCE/DOCKET NUMBER: GI6000
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 498-8224
 ; TELEFAX: (617) 876-5851

```
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 470 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
; US-08-833-823-2

Query Match      14.9%; Score 702; DB 2; Length 470;
Best Local Similarity 39.2%; Pred. No. 7.5e-54;
Matches 177; Conservative 87; Mismatches 161; Indels 26; Gaps 8;

QY 5 LLPVS-LLSVAVSAIKELPGVKYVYVYVIRLPHLEKREAKEPE-----QOEQWETELK 58
Db 18 LLPVLWIVQTAIAIKQTPELTILHEIVCPKKLHLHREIKNNQTEKHGKEERYEPEVQ 77
QY 59 YKMTINGKIAVLYLKKNKLLAPGYTETYNSTGKEITTSQIMWVYVYOGHILNEKYSW 118
Db 78 YQMILNGEEIILSLQKTKHLLGPDYETELYSRGEIITKPEMHECHYKGNILNEKNSV 137
QY 119 ASISTWGLRGYMQGQRYWIEPLSPHRRGQEHALWKYPWEKN-----YWSWGM 171
Db 138 ASISTCDGLRGYFTHHQRYQIKPLKSTDE--KEHAVETSQEEQDPANHTCGVKSTDGK 195
QY 172 WGVLAHWLQONIALPATKLVKLKWRKQVEHEKIEYVYLWNGEWKRYNENQWEIRKV 231
Db 196 QGPI---RISRLKSPE---KEDFLRAQ---KYIDLVLVDNAFYKYNENLILRSFV 245
QY 232 WEMANYNMLYKLLNTHVALVGMIEWTWKIKITPNASWTLENKSWKRGSVLSRRKHWH 291
Db 246 FVNNLLNVIYNTIDVQVALVGMIEWSDGDKIKVVPASITFDNPLRWHSSNLG-KKIH 304
QY 292 IAQLITATELAGTTVGLAWMTSPYSVGVVQVHSMNLLRVAGTMAHEMCHNMGWHHW 351
Db 305 HAQLLSGISFNRRVGLAASNSLSPSSVAIVAEAKKONVALGVMSHELGHVLMGPDVP 364
QY 352 YSKWKPSTIWMKALSNYIPTWSSWSRLSYKWKWENKLSNWLWNAPLTWIISTPIWG 411
Db 365 FNTKCPSSGCVMMNQYLSKFKPFDFSTSCRAHFERYLLSQKPKCLLOAPINTMPTVCG 424
QY 412 NOLVEMGEWMMGTSEWTNIWNAKWIK 442
Db 425 NHLEVEDCDGSGPKECTNLCCALCKLK 455

RESULT 13
US-08-833-150-2
; Sequence 2, Application US/08813150
; Patent No. 6069229
; GENERAL INFORMATION:
; APPLICANT: Mueller, Christopher
; APPLICANT: Lebecque, Serge J.E.
; APPLICANT: Liu, Yong-Jun
; APPLICANT: Dowling, Lynette M.
; APPLICANT: Huffine, Constance M.
; APPLICANT: Gorman, Daniel M.
; TITLE OF INVENTION: MAMMALIAN PROTEINASES; OXIDOREDUCTASES;
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/813,150
```

```
; FILING DATE: 07-MAR-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: SP0693
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-852-9196
; TELEFAX: 650-496-1200
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 470 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
; US-08-813-150-2

Query Match      14.6%; Score 690; DB 3; Length 470;
Best Local Similarity 38.8%; Pred. No. 8.6e-53;
Matches 175; Conservative 88; Mismatches 162; Indels 26; Gaps 8;

QY 5 LLPVS-LLSVAVSAIKELPGVKYVYVYVIRLPHLEKREAKEPE-----QOEQWETELK 58
Db 18 LLPVLWIVQTAIAIKQTPELTILHEIVCPKKLHLHREIKNNQTEKHGKEERYEPEVQ 77
QY 59 YKMTINGKIAVLYLKKNKLLAPGYTETYNSTGKEITTSQIMWVYVYOGHILNEKYSW 118
Db 78 YQMILNGEEIILSLQKTKHLLGPDYETELYSRGEIITKPEMHECHYKGNILNEKNSV 137
QY 119 ASISTWGLRGYMQGQRYWIEPLSPHRRGQEHALWKYPWEKN-----YWSWGM 171
Db 138 ASISTCDGLRGYFTHHQRYQIKPLKSTDE--KEHAVETSQEEQDPANHTCGVKSTDGK 195
QY 172 WGVLAHWLQONIALPATKLVKLKWRKQVEHEKIEYVYLWNGEWKRYNENQWEIRKV 231
Db 196 QGPI---RISRLKSPE---KEDFLRAQ---KYIDLVLVDNAFYKYNENLILRSFV 245
QY 232 WEMANYNMLYKLLNTHVALVGMIEWTWKIKITPNASWTLENKSWKRGSVLSRRKHWH 291
Db 246 FVNNLLNVIYNTIDVQVALVGMIEWSDGDKIKVVPASITFDNPLRWHSSNLG-KKIH 304
QY 292 IAQLITATELAGTTVGLAWMTSPYSVGVVQVHSMNLLRVAGTMAHEMCHNMGWHHW 351
Db 305 HAQLLSGISFNRRVGLAASNSLSPSSVAIVAEAKKONVALGVMSHELGHVLMGPDVP 364
QY 352 YSKWKPSTIWMKALSNYIPTWSSWSRLSYKWKWENKLSNWLWNAPLTWIISTPIWG 411
Db 365 FNTKCPSSGCVMMNQYLSKFKPFDFSTSCRAHFERYLLSQKPKCLLOAPINTMPTVCG 424
QY 412 NOLVEMGEWMMGTSEWTNIWNAKWIK 442
Db 425 NHLEVEDCDGSGPKECTNLCCALCKLK 455

RESULT 14
US-09-546-553-2
; Sequence 2, Application US/09546553
; Patent No. 6518405
; GENERAL INFORMATION:
; APPLICANT: Mueller, Christopher
; APPLICANT: Lebecque, Serge J.E.
; APPLICANT: Liu, Yong-Jun
; APPLICANT: Dowling, Lynette M.
; APPLICANT: Huffine, Constance M.
; APPLICANT: Gorman, Daniel M.
; TITLE OF INVENTION: MAMMALIAN PROTEINASES; OXIDOREDUCTASES;
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
```

```
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/546,553
; FILING DATE: 10-APR-2000
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/813,150
; FILING DATE: 07-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: SF0693
; TELEPHONE: 650-852-9196
; TELEFAX: 650-496-1200
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 470 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-546-553-2

Query Match 14.6%; Score 690; DB 4; Length 470;
Best Local Similarity 38.6%; Pred. No. 8,6e-53;
Matches 175; Conservative 88; Mismatches 162; Indels 26; Gaps 8;

QY 5 LLPVS-LLLSVAVSAIKELPGVKYEVVYPIRLHLKRAKEPE-----QOEQWETELK 58
Db 18 LLPVLWLVQTOAIAIKOTPELTLEIVCPKXHLHLKRAEIKNNQTEKHGEERYEPEVQ 77

QY 59 YKMTINGIAVLVYKKNKLLAPGYTETVYNTSGKEITTSPOIMWWWYQGHILNEKVS 118
Db 78 YQMLNGEIEILSLQTKHLLGFDTELDYSPRGEITTKPENMEHCYKGNILNEKVS 137

QY 119 ASISTWRGLRGYSQGWORYMTEPLSPIHRWQOEHALWKYNPWEKN-----YASTGCM 171
Db 138 ASISTCDGLRGYFTHHHQRYQIKPLKSTDE--KEHAFVTSNQEODPANHTCGVASTGK 195

QY 172 WGVLMWHLQNNIALPATKVLKRWKXQEHKYEYVYLVNNGEKRYNENQWBIKRV 231
Db 196 QGPI---RISRLKSPE---KEDELRAQ---KYIDLVLVDNAFYKNYNENLTIRSFV 245

QY 232 WEMANYVMYKKNLTHVALVGMETWKNKIKITENASWTLENKSKRGSVLSRKRM 291
Db 246 FVVMNLLVYNTIDVQVALVGMETWSQGEKIKVFSASTTFDNFLRHSS--NPGKKLHD 304

QY 292 IAQLTATLACTVGLAMSTMSPYSVGVVQVHWSNLLRVAGTMAHEMGHNGWGHWM 351
Db 305 HAQLLSGLISFNRRVGLAASNSLSPSSVAVTEAKKNVALVGVMSHELGHVGMPEVP 364

QY 352 YSWKPPSTIWMWKAISWIPTWNSWSRLSTWKKWELKSNLWLNAPLPTWIISTPTWG 411
Db 365 FNTKPESSCVNQYLSKFPKDFSTCSRAHERYLLSKOKPCLLQAPITNIMTTPVG 424

QY 412 NCLVEMGMMWMTSEETNIWMWAKTKIK 442
Db 425 NHLLEVGEDCDGSPKECTSLCCALTKIK 455

RESULT 15
US-09-026-001A-14
; Sequence 14, Application US/09026001A
; Patent No. 6413760
; GENERAL INFORMATION:
; APPLICANT: Boodhoo, Amechand
; APPLICANT: Seehra, Jasbir
```

```
; APPLICANT: Shaw, Gray
; APPLICANT: Sako, Dianne
; TITLE OF INVENTION: HIGHLY PURIFIED MOCARHAGIN, A COBRA VENOM
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/026,001A
; FILING DATE: 18-FEB-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G152938
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 592 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-026-001A-14

Query Match 14.3%; Score 674; DB 4; Length 592;
Best Local Similarity 32.1%; Pred. No. 3e-51;
Matches 202; Conservative 98; Mismatches 287; Indels 42; Gaps 10;

QY 1 MQQGLPLVSLLSV---AVSAIKELPGVKYEVVYPIRLHLKRAKEPEQOEQWETEL 57
Db 1 MQQALLVIICLVVFPYQGSIIILSGNVNDYEVVYQKVPALLKGGVQNPQETKYEDTM 60

QY 58 KYMTINGIAVLVYKKNKLLAPGYTETVYNTSGKEITTSPOIMWWWYQGHILNEKVS 117
Db 61 QYEFQVNGEPVVLHLERNKGLFSEDYETHTYAPDGRREITTSPPVQDCHCYHYIQNEADS 120

QY 118 WASISTWRGLRGYSQGWORYMTEPLSPIHRWQOEHALWKYNPWEKNYSTWGMGVLWA 177
Db 121 SAIIISACDLKGFHKHQQGETYFIEPLKLPD--SESHAIYKDNVB--NEDETPTCGTET 177

QY 178 HMLQONIALPATKVLKRWK--VQBEKYEYVYLVNNGEKRYNENQWBIKRVWEMAN 236
Db 178 TWESDESIEKTSQTLNTPQDGVLAQKYEYFVVVDNRMYRYKNEBPAIKERVYEMVN 237

QY 237 YNNMLYKKNLTHVALVGMETWKNKIKITENASWTLENKSKRGSVLSRKHNAQLI 296
Db 238 AVNTYRPLKIHITLIGLEINSGDDKFEVYKPVAGATLKSFRDWRVETVLLPRKENDNAQL 297

QY 297 TATELAGTTVGLAMSTMSPYSVGVVQVHWSNLLRVAGTMAHEMGHNGWGHWMYSWKW 356
Db 298 TGDIFNGTVVGYATYGTTLCTQNSVAVVQDYNRKISNVASTMAHELGHNILGHDGASCIC 357

QY 357 PSTIWMWKAISWIPTW--WSSWSRLSTWKKWELKSNLWLNAPLPTWIISTPTWG 415
Db 358 SLRPLCMSKGR--APAFQFSSCVREYREYLLRRPQCILNKPLSTDTVSPAICNGYFV 415

QY 416 EMGEWMMWMTSEETNIWMWAKTKIKATQWALGEMWKEKWKQKAGMWRPRAKWEWLP 475
Db 416 ERGEEDCDGSPADCCSACCDAATCKFK-----GGEAR-----CRAAKDDCDLP 458
```


Qy	476	EMWNGKSGNWPRWQVNGWPHHGGHLMGTWETLOEQWTELWGPCTEVAVKSWYHRN	535
Db	459	ELCTERSVECFDLSLQRNGHPCQNNKGYCYNGACFTTNQCIALMGTDFTVSPDGCFLN	518
Qy	536	EGGSKYGYWRRVWNTLIEWKANWTWNGKLTWQGGSWNLFWKGRIVTWLTWKTWPEWT SQ	595
Db	519	VRGNDVSHCRKENGAKIPCAAADVKGRLYCTERD-----TMSCRPP--LDP	563
Qy	596	EIGMVANGTKWGNKYWINAEWWIEKAY	624
Db	564	DGVNAEPGTCGDGMVCSNGOCVNVQTAY	592

Search completed: June 9, 2004, 18:03:28
Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2004, 18:02:26 ; Search time 48 Seconds
(without alignments)
4542.447 Million cell updates/sec

Title: US-10-726-148A-15
Perfect score: 4712
Sequence: 1 MLOGLPVSLLSVAVSAIK.....PPTVAKWNPMTPKNSNPKA 775

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:

```
1: /cgm2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgm2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgm2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgm2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgm2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgm2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgm2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgm2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgm2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgm2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgm2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgm2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgm2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgm2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgm2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgm2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgm2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgm2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3128	66.4	788	12	US-10-726-774-2299
2	1610	34.2	439	14	US-10-106-698-4694
3	1635	34.1	435	9	US-09-955-504-11
4	1605	34.1	435	13	US-10-125-470-11
5	1605	34.1	435	13	US-10-125-452-11
6	854	18.1	233	9	US-09-925-299-1014
7	854	18.1	233	10	US-09-925-299-1014
8	815	17.3	208	9	US-09-955-504-12
9	815	17.3	208	13	US-10-125-470-12
10	815	17.3	208	13	US-10-125-452-12
11	773.5	16.4	776	9	US-09-908-193-33
12	764.5	16.2	788	9	US-09-908-193-35
13	755.5	16.0	611	9	US-09-921-823-17
14	755	16.0	775	12	US-10-232-972B-2
15	750.5	15.9	779	9	US-09-908-193-10

16	750	15.9	754	9	US-09-908-193-32	Sequence 32, Appl
17	745	15.8	778	9	US-09-908-193-8	Sequence 8, Appl
18	736	15.6	695	9	US-09-764-898-191	Sequence 191, Appl
19	735.5	15.6	613	9	US-09-996-620-10	Sequence 10, Appl
20	732.5	15.5	621	9	US-09-996-620-18	Sequence 18, Appl
21	717	15.2	621	9	US-09-996-620-6	Sequence 6, Appl
22	702	14.9	470	13	US-10-114-891-8	Sequence 8, Appl
23	690	14.6	470	14	US-10-349-806-2	Sequence 2, Appl
24	685	14.5	606	14	US-10-439-532-12	Sequence 12, Appl
25	674	14.3	592	9	US-09-996-620-14	Sequence 14, Appl
26	662	14.0	571	14	US-10-439-532-13	Sequence 13, Appl
27	647	13.7	521	9	US-09-996-620-12	Sequence 12, Appl
28	642	13.6	824	12	US-10-211-858-58	Sequence 58, Appl
29	642	13.6	824	14	US-10-226-844-1	Sequence 1, Appl
30	642	13.6	824	14	US-10-210-951-58	Sequence 58, Appl
31	642	13.6	824	14	US-10-211-884-58	Sequence 58, Appl
32	596	12.6	439	9	US-09-996-620-8	Sequence 8, Appl
33	586.5	12.4	483	14	US-10-439-532-2	Sequence 2, Appl
34	543	11.5	434	9	US-09-921-823-20	Sequence 20, Appl
35	539	11.4	918	9	US-09-955-504-9	Sequence 9, Appl
36	539	11.4	918	13	US-10-020-733-4	Sequence 4, Appl
37	539	11.4	918	13	US-10-125-470-9	Sequence 9, Appl
38	539	11.4	918	13	US-10-125-452-9	Sequence 9, Appl
39	539	11.4	918	15	US-10-274-639-18	Sequence 18, Appl
40	539	11.4	918	16	US-10-333-574-18	Sequence 18, Appl
41	539	11.4	955	13	US-10-020-733-8	Sequence 8, Appl
42	538.5	11.4	956	16	US-10-408-765A-2367	Sequence 2367, Ap
43	536	11.4	481	16	US-10-383-588A-8	Sequence 8, Appl
44	535	11.4	926	13	US-10-020-733-2	Sequence 2, Appl
45	535	11.4	963	13	US-10-020-733-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1

```
US-10-276-774-2299
: Sequence 2299, Application US/10276774
: Publication No. US20040053245A1
: GENERAL INFORMATION:
: APPLICANT: Hyseq, Inc.
: TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
: FILE REFERENCE: 21272-030
: CURRENT APPLICATION NUMBER: US/10/276, 774
: CURRENT FILING DATE: 2002-11-18
: PRIOR APPLICATION NUMBER: 09/560, 875
: PRIOR FILING DATE: 2000-04-27
: PRIOR APPLICATION NUMBER: 09/496, 914
: PRIOR FILING DATE: 2000-02-03
: NUMBER OF SEQ ID NOS: 2700
: SOFTWARE: Custom
: SEQ ID NO 2299
: LENGTH: 788
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-276-774-2299
```

Query Match 66.4%; Score 3128; DB 12; Length 788;
Best Local Similarity 84.4%; Pred. No. 3.7e-242;
Matches 654; Conservative 29; Mismatches 92; Indels 0; Gaps 0;

QY	1	MLOGLPVSLLSVAVSAIKELPGVKKEVVPYIRLPHLHKREAKEPEQOEOMTELKVK	60
DB	14	MLOGLPVSLLSVAVSAIKELPGVKKEVVPYIRLPHLHKREAKEPEQOEOMTELKVK	73
QY	61	MTINGKIAVLKKNKNLLAPGYTYNSTGKEITTSPOIMWYVYQGHILNEKYSWAS	120
DB	74	MTINGKIAVLKKNKNLLAPGYTYNSTGKEITTSPOIMWYVYQGHILNEKYSWAS	133
QY	121	ISTWRLRGYVSGQWRYWIBPLSPHHRMGQSHALMKYKPNWKNYSTWGMGVLMAHML	180
DB	134	ISTCRLRGYVSGQWRYWIBPLSPHHRMGQSHALMKYKPNWKNYSTWGMGVLMAHML	193


```
Db 121 AGMVCRAKDECDLPEVCMNGKSGNCPDDPQVNGPFCCHGKGRCLMGTCTLOEQCTELW 180
Qy 521 GEGTEVAKWSYNNRNEGSGKYGYMRRVWMTLLPEKANKWTTWGLKAWQGGSNLPMKGRIV 580
Db 181 GGEVEADKSCYNNRNEGSGKYGYCRRVDTLLPKRANTMCGKLFQCGGSDNLPWKGRIV 240
Qy 581 TWLTKTWPEWTSQIEGMVANGTKWGNKVINABWVIEKAYKSTNWSKWKGHAVW 640
Db 241 TFLTKTFDPEDTSQIEGMVANGTKGDNKVCINAECDVIEKAYKSTNCSKCKGHAVCD 300
Qy 641 HELQWQEBGWIPTPWSSVWHSIVVGVLPMAVIVVWVAMVIRHQSREKQKQW 700
Db 301 HELQWQEBGWIPTPDCCDSSVVFHFSIVVGVLPMAVIFVWVAMVIRHQSREKQKQD 360
Qy 701 PLSTTGTRPHKQKRPQWYKAVQPMQSKMHPHYVLPVEGNEPPASPHKDTNALPPTVW 760
Db 361 PLSTTGTRPHKQKRPQWYKAVQPMQSKMHPHYVLPVEGNEPPASPHKDTNALPPTV 420
Qy 761 KWNPMSTPKWSNPKA 775
Db 421 KWNPMSTPKDSNPKA 435

RESULT 4
US-10-125-470-11
; Sequence 11, Application US/10125470
; Publication No. US20020165377A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: ADAM Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PTO66P1
; CURRENT APPLICATION NUMBER: US/10/125,470
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US/09/712,907A
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: 60/136,388
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: 60/142,930
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 60/178,717
; PRIOR FILING DATE: 2000-01-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-125-470-11

Query Match 34.1%; Score 1605; DB 13; Length 435;
Best Local Similarity 79.1%; Pred. No. 2.5e-120;
Matches 344; Conservative 20; Mismatches 71; Indels 0; Gaps 0;

Qy 341 MGHNMGWMTWYYSWKWPSTIIVVWVKALSWIPTWSSWSRLSYWKWWEKLSNLMNAPL 400
Db 1 MGHNFGHDDYSCKCPSTICVMDKALSFYIPTDFSSCSRLSYDKFEDKLSNCLFNAPL 60
Qy 401 PTWIISTPIGNQLVEMGEMWMMGTSEETNINWNAKTWIKIKATWQALGEMWKEWQWK 460
Db 61 PTDIISTPICGNQLVEMGEDCDGCTSEECTNICCDAKTCKIKATFQCALGECCEKQFK 120
Qy 461 AGWVPAKWEWLPEWNNCKSGNWPWQVNGWPHHGKGLMGTWTLQOWTELW 520
Db 121 AGMVCRAKDECDLPEMCMNGKSGNCPDDPQVNGPFCCHGKGRCLMGTCTLOEQCTELW 180
Qy 521 GPGEVAKWSYNNRNEGSGKYGYMRRVWMTLLPEKANKWTTWGLKAWQGGSNLPMKGRIV 580
Db 181 GPGEVADKSCYNNRNEGSGKYGYCRRVDTLLPKRANTMCGKLFQCGGSDNLPWKGRIV 240
Qy 581 TWLTKTWPEWTSQIEGMVANGTKWGNKVINABWVIEKAYKSTNWSKWKGHAVW 640
Db 241 TFLTKTFDPEDTSQIEGMVANGTKGDNKVCINAECDVIEKAYKSTNCSKCKGHAVCD 300
Qy 641 HELQWQEBGWIPTPWSSVWHSIVVGVLPMAVIVVWVAMVIRHQSREKQKQW 700
Db 301 HELQWQEBGWIPTPDCCDSSVVFHFSIVVGVLPMAVIFVWVAMVIRHQSREKQKQD 360
Qy 701 PLSTTGTRPHKQKRPQWYKAVQPMQSKMHPHYVLPVEGNEPPASPHKDTNALPPTVW 760
Db 361 PLSTTGTRPHKQKRPQWYKAVQPMQSKMHPHYVLPVEGNEPPASPHKDTNALPPTV 420
```

```
Qy 641 HELQWQEBGWIPTPWSSVWHSIVVGVLPMAVIVVWVAMVIRHQSREKQKQW 700
Db 301 HELQWQEBGWIPTPDCCDSSVVFHFSIVVGVLPMAVIFVWVAMVIRHQSREKQKQD 360
Qy 701 PLSTTGTRPHKQKRPQWYKAVQPMQSKMHPHYVLPVEGNEPPASPHKDTNALPPTVW 760
Db 361 PLSTTGTRPHKQKRPQWYKAVQPMQSKMHPHYVLPVEGNEPPASPHKDTNALPPTV 420
Qy 761 KWNPMSTPKWSNPKA 775
Db 421 KWNPMSTPKDSNPKA 435

RESULT 5
US-10-125-452-11
; Sequence 11, Application US/10125452
; Publication No. US20020173640A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: ADAM Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PTO66P2
; CURRENT APPLICATION NUMBER: US/10/125,452
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: 09/955,504
; PRIOR FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: 09/712,907
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: PCT/US00/14308
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/178,717
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/142,930
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 60/136,388
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-125-452-11

Query Match 34.1%; Score 1605; DB 13; Length 435;
Best Local Similarity 79.1%; Pred. No. 2.5e-120;
Matches 344; Conservative 20; Mismatches 71; Indels 0; Gaps 0;

Qy 341 MGHNMGWMTWYYSWKWPSTIIVVWVKALSWIPTWSSWSRLSYWKWWEKLSNLMNAPL 400
Db 1 MGHNFGHDDYSCKCPSTICVMDKALSFYIPTDFSSCSRLSYDKFEDKLSNCLFNAPL 60
Qy 401 PTWIISTPIGNQLVEMGEMWMMGTSEETNINWNAKTWIKIKATWQALGEMWKEWQWK 460
Db 61 PTDIISTPICGNQLVEMGEDCDGCTSEECTNICCDAKTCKIKATFQCALGECCEKQFK 120
Qy 461 AGWVPAKWEWLPEWNNCKSGNWPWQVNGWPHHGKGLMGTWTLQOWTELW 520
Db 121 AGMVCRAKDECDLPEMCMNGKSGNCPDDPQVNGPFCCHGKGRCLMGTCTLOEQCTELW 180
Qy 521 GPGEVAKWSYNNRNEGSGKYGYMRRVWMTLLPEKANKWTTWGLKAWQGGSNLPMKGRIV 580
Db 181 GPGEVADKSCYNNRNEGSGKYGYCRRVDTLLPKRANTMCGKLFQCGGSDNLPWKGRIV 240
Qy 581 TWLTKTWPEWTSQIEGMVANGTKWGNKVINABWVIEKAYKSTNWSKWKGHAVW 640
Db 241 TFLTKTFDPEDTSQIEGMVANGTKGDNKVCINAECDVIEKAYKSTNCSKCKGHAVCD 300
Qy 641 HELQWQEBGWIPTPWSSVWHSIVVGVLPMAVIVVWVAMVIRHQSREKQKQW 700
Db 301 HELQWQEBGWIPTPDCCDSSVVFHFSIVVGVLPMAVIFVWVAMVIRHQSREKQKQD 360
Qy 701 PLSTTGTRPHKQKRPQWYKAVQPMQSKMHPHYVLPVEGNEPPASPHKDTNALPPTVW 760
```

Db 361 PLSTGTRPHKQKQPMVAVQPMKPHVYDLPEVNEPPASFKDNLPTVVF 420
Qy 761 KWNPMSTPKNSNPKA 775
Db 421 KDNEMSTPKDSNPKA 435

RESULT 6

US-09-925-299-1014
; Sequence 1014, Application US/09925299
; Patent No. US20020355627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PAL02
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1014
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (44)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (56)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (71)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1014

Query Match 18.1%; Score 854; DB 9; Length 233;

Best Local Similarity 82.5%; Pred. No. 2e-60; Indels 0; Gaps 0;
Matches 179; Conservative 9; Mismatches 29;

Qy 559 TWWGKLMWQGGSNLPPWKGRIVTTLTKTWPEWTSQBIGMVANGTKWGNKVNINAEWV 618
Db 17 TWCGLFCQGGSDNLPWKGRIVTFLTCXTPPEDTSQELXVANGTKCGDNKVCNAECV 76
Qy 619 WIEKAYKSTWSSSKWKGHAVWVHLEQWEGWIPPPWWWSSVWVHWSIVVGVLPMAVI 678
Db 77 DIEKAYKSTNCSKCKGHAVCDHELCQCEEGWIPDDSDSSVVFHFSIVVGVLPMAVI 136
Qy 679 WVVAVMVRHOSREKOKKORPLSTTGTTRPHKQKQPMVAVQPMKPHVYDLPE 738
Db 137 FVVAVMVRHOSREKOKKORPLSTTGTTRPHKQKQPMVAVQPMKPHVYDLPE 196
Qy 739 VEGNEPPASFKDNLPTVTKWNPMTPKNSNPKA 775
Db 197 VEGNEPPASFKDNLPTVTKWNPMTPKNSNPKA 233

RESULT 7

US-09-925-299-1014
; Sequence 1014, Application US/09925299
; Patent No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PAL02
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1014
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (44)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (56)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (71)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1014

Query Match 18.1%; Score 854; DB 10; Length 233;

Best Local Similarity 82.5%; Pred. No. 2e-60; Indels 0; Gaps 0;
Matches 179; Conservative 9; Mismatches 29;

Qy 559 TWWGKLMWQGGSNLPPWKGRIVTTLTKTWPEWTSQBIGMVANGTKWGNKVNINAEWV 618
Db 17 TWCGLFCQGGSDNLPWKGRIVTFLTCXTPPEDTSQELXVANGTKCGDNKVCNAECV 76
Qy 619 WIEKAYKSTWSSSKWKGHAVWVHLEQWEGWIPPPWWWSSVWVHWSIVVGVLPMAVI 678
Db 77 DIEKAYKSTNCSKCKGHAVCDHELCQCEEGWIPDDSDSSVVFHFSIVVGVLPMAVI 136
Qy 679 WVVAVMVRHOSREKOKKORPLSTTGTTRPHKQKQPMVAVQPMKPHVYDLPE 738
Db 137 FVVAVMVRHOSREKOKKORPLSTTGTTRPHKQKQPMVAVQPMKPHVYDLPE 196
Qy 739 VEGNEPPASFKDNLPTVTKWNPMTPKNSNPKA 775
Db 197 VEGNEPPASFKDNLPTVTKWNPMTPKNSNPKA 233

RESULT 8

US-09-955-504-12
; Sequence 12, Application US/09955504
; Publication No. US20020182702A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.

; TITLE OF INVENTION: ADAM Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PTO0692
; CURRENT APPLICATION NUMBER: US/09/955,504
; CURRENT FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: 60/234,222
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: 09/712,907
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: PCT/US00/14308
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/178,717
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/142,930
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 60/136,388
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 12
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-955-504-12

Query Match 17.3%; Score 815; DB 9; Length 208;
Best Local Similarity 80.1%; Pred. No. 2.4e-57;

```
Matches 173; Conservative 9; Mismatches 26; Indels 8; Gaps 1;
QY 560 MWGLWMOGGSNLPWKGRIVTWLTKTWPEWTSQEIOMVANGTKGWNKVINAEWVW 619
Db 1 MCGKLFQCGSSDNLFPWKGRIVTFLTKTFDPEDTSQEIOMVANGTKGDNKVCINAEVD 60
QY 620 IEKAYKSTNWSKKGHAYVWHELOQWEEGIPPPWVWSSVWVHWSIVVGLWPMVAVI 679
Db 61 IEKAYKSTNCSKKGHAYVWHELOQWEEGIPPPWVWSSVWVHWSIVVGLWPMVAVI 120
QY 680 VVAVVIRHQSREKQKQWQRLSTGTTPHKKQKQKQVQPMKAVQPOEMSKPHVYVLPV 739
Db 121 VVAVVIRHQSREKQKQVQRLSTGTTPHKKQKQKQPM-----MSOMKPHVYDLPV 172
QY 740 EGNEPPASHHKWTNALPPTVFWKNNPMSTPKNSPKA 775
Db 173 EGNEPPASPHKDTNALPPTVFWKNNPMSTPKNSPKA 208

RESULT 9
US-10-125-470-12
; Sequence 12, Application US/10125470
; Publication No. US20020165377A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: ADAM Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT006P1
; CURRENT APPLICATION NUMBER: US/10/125,470
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US/09/712,907A
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: 60/136,388
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: 60/142,930
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 60/178,717
; PRIOR FILING DATE: 2000-01-28
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-125-470-12

Query Match 17.3%; Score 815; DB 13; Length 208;
Best Local Similarity 80.1%; Pred. No. 2.4e-57;
Matches 173; Conservative 9; Mismatches 26; Indels 8; Gaps 1;
QY 560 MWGLWMOGGSNLPWKGRIVTWLTKTWPEWTSQEIOMVANGTKGWNKVINAEWVW 619
Db 1 MCGKLFQCGSSDNLFPWKGRIVTFLTKTFDPEDTSQEIOMVANGTKGDNKVCINAEVD 60
QY 620 IEKAYKSTNWSKKGHAYVWHELOQWEEGIPPPWVWSSVWVHWSIVVGLWPMVAVI 679
Db 61 IEKAYKSTNCSKKGHAYVWHELOQWEEGIPPPWVWSSVWVHWSIVVGLWPMVAVI 120
QY 680 VVAVVIRHQSREKQKQWQRLSTGTTPHKKQKQKQVQPMKAVQPOEMSKPHVYVLPV 739
Db 121 VVAVVIRHQSREKQKQVQRLSTGTTPHKKQKQKQPM-----MSOMKPHVYDLPV 172
QY 740 EGNEPPASHHKWTNALPPTVFWKNNPMSTPKNSPKA 775
Db 173 EGNEPPASPHKDTNALPPTVFWKNNPMSTPKNSPKA 208

RESULT 10
US-10-125-452-12
; Sequence 12, Application US/10125452
; Publication No. US20020173640A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: ADAM Polynucleotides, Polypeptides, and Antibodies
```

```
FILE REFERENCE: PT006P2
; CURRENT APPLICATION NUMBER: US/10/125,452
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: 09/955,504
; PRIOR FILING DATE: 2001-09-19
; PRIOR APPLICATION NUMBER: 09/712,907
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: PCT/US00/14308
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/178,717
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/142,930
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 60/136,388
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-125-452-12

Query Match 17.3%; Score 815; DB 13; Length 208;
Best Local Similarity 80.1%; Pred. No. 2.4e-57;
Matches 173; Conservative 9; Mismatches 26; Indels 8; Gaps 1;
QY 560 MWGLWMOGGSNLPWKGRIVTWLTKTWPEWTSQEIOMVANGTKGWNKVINAEWVW 619
Db 1 MCGKLFQCGSSDNLFPWKGRIVTFLTKTFDPEDTSQEIOMVANGTKGDNKVCINAEVD 60
QY 620 IEKAYKSTNWSKKGHAYVWHELOQWEEGIPPPWVWSSVWVHWSIVVGLWPMVAVI 679
Db 61 IEKAYKSTNCSKKGHAYVWHELOQWEEGIPPPWVWSSVWVHWSIVVGLWPMVAVI 120
QY 680 VVAVVIRHQSREKQKQWQRLSTGTTPHKKQKQKQVQPMKAVQPOEMSKPHVYVLPV 739
Db 121 VVAVVIRHQSREKQKQVQRLSTGTTPHKKQKQKQPM-----MSOMKPHVYDLPV 172
QY 740 EGNEPPASHHKWTNALPPTVFWKNNPMSTPKNSPKA 775
Db 173 EGNEPPASPHKDTNALPPTVFWKNNPMSTPKNSPKA 208

RESULT 11
US-09-908-193-33
; Sequence 33, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: SHIMKETS, RICHARD A.
; APPLICANT: ZERHUSEN, BRYAN
; APPLICANT: MALYANKAR, URIEL M.
; APPLICANT: PADIGARU, MURALIDHARA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY
; FILE REFERENCE: 21402-062
; CURRENT APPLICATION NUMBER: US/09/908,193
; CURRENT FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: 60/220,273
; PRIOR FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: 60/221,650
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 60/221,233
; PRIOR FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 60/220,912
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/218,875
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,870
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: 60/218,901
; PRIOR FILING DATE: 2000-07-18
; NUMBER OF SEQ ID NOS: 74
```

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 776
; TYPE: PRT
; ORGANISM: Macaca fascicularis
US-09-908-193-33

Query Match      16.4%; Score 773.5; DB 9; Length 776;
Best Local Similarity 30.8%; Pred. No. 2.3e-53;
Matches 247; Conservative 137; Mismatches 322; Indels 97; Gaps 20;

Qy 1 MLOGLPVSLLSVAVSAIKELPGVKYEVVYPIRLPHLKE--AKEPEQOQOMETEL 57
Db 1 MLOGLPVSLLSVAVSAIKELPGVKYEVVYPIRLPHLKE--AKEPEQOQOMETEL 57
Qy 58 KYMTNGKIAVLTKKKNLLAPGYTETYNSTGKEITTSPTQIMWVYQGHILNEKVS 117
Db 58 KYMTNGKIAVLTKKKNLLAPGYTETYNSTGKEITTSPTQIMWVYQGHILNEKVS 117
Qy 60 MYBIKLRKTLVLHLLRSREFLGSNYSETFYSMKGEAFTRHLQIMDCHCFYQGSIVHEYDS 119
Db 60 MYBIKLRKTLVLHLLRSREFLGSNYSETFYSMKGEAFTRHLQIMDCHCFYQGSIVHEYDS 119
Qy 118 WASISTWGLRGVWSQGWRYWIEPLSPIHRWGQEHALWKY---PWEKYNWSTWGMGV 174
Db 118 WASISTWGLRGVWSQGWRYWIEPLSPIHRWGQEHALWKY---PWEKYNWSTWGMGV 174
Qy 120 AASISTCKLGRGFRVNDORYLIEPVK--YSDEGEHLVFKYNPVPVYVANSCT----- 172
Db 120 AASISTCKLGRGFRVNDORYLIEPVK--YSDEGEHLVFKYNPVPVYVANSCT----- 172
Qy 175 LNAHWLQONIAL-----PATKLVKLKWRKVQEHKYEYLVVLWNGSKRYNE 222
Db 175 LNAHWLQONIAL-----PATKLVKLKWRKVQEHKYEYLVVLWNGSKRYNE 222
Qy 172 -----ELNFRKTVPGDTESEGDPKMKAI-----HNEKYLELFIVADDTVYRRNSH 217
Db 172 -----ELNFRKTVPGDTESEGDPKMKAI-----HNEKYLELFIVADDTVYRRNSH 217
Qy 223 NOWEIRKRVEMANYVMYKKNLTHVALGMEIWTWKKIKITPNASWTLENMSKWRGS 282
Db 223 NOWEIRKRVEMANYVMYKKNLTHVALGMEIWTWKKIKITPNASWTLENMSKWRGS 282
Qy 218 PHNKLNRRIQGVNFVNIYKTLNIHTLVGIEIWTHEDKIELHSNIETTLRFSSQWER 277
Db 218 PHNKLNRRIQGVNFVNIYKTLNIHTLVGIEIWTHEDKIELHSNIETTLRFSSQWER 277
Qy 283 VLSRRKHWHIAQLTATLHAGTIVGLAWMSTWSP--YSVGUVVQVHNSNLLR-----VAGTM 337
Db 283 VLSRRKHWHIAQLTATLHAGTIVGLAWMSTWSP--YSVGUVVQVHNSNLLR-----VAGTM 337
Qy 278 ILKTRKDFHVLVLSGKWIYTHVQGISYPAGMCLPYSTSIK-----DLLPDTNIIANRM 333
Db 278 ILKTRKDFHVLVLSGKWIYTHVQGISYPAGMCLPYSTSIK-----DLLPDTNIIANRM 333
Qy 338 AHENGHNGWHWYKWSKVESTIWMKALSWYIPTW--WSSWSRLSYKWKWENKLSNWLW 396
Db 338 AHENGHNGWHWYKWSKVESTIWMKALSWYIPTW--WSSWSRLSYKWKWENKLSNWLW 396
Qy 334 AHQIGHNLGQWDEFPCTPSGKCVMSDGS--IPALKFSKCSQVQHYLYKDKPTCML 391
Db 334 AHQIGHNLGQWDEFPCTPSGKCVMSDGS--IPALKFSKCSQVQHYLYKDKPTCML 391
Qy 397 NAFPLTWIISTPIMGNOLVEMGEWWWGTSEBTNIWWAKTWKIKATWQWALGEMWKR 456
Db 397 NAFPLTWIISTPIMGNOLVEMGEWWWGTSEBTNIWWAKTWKIKATWQWALGEMWKR 456
Qy 392 NIIPPONFDDPFCGKGLDBEBCDCGSPQECTNPCCDAHTCVLKPGETCAEGCECSC 451
Db 392 NIIPPONFDDPFCGKGLDBEBCDCGSPQECTNPCCDAHTCVLKPGETCAEGCECSC 451
Qy 457 QWKKAGYVWPAKWEWMLPEMNGKSGNNEPWRQVNGWPHHKGHGLAGTWTPLQEQW 516
Db 457 QWKKAGYVWPAKWEWMLPEMNGKSGNNEPWRQVNGWPHHKGHGLAGTWTPLQEQW 516
Qy 452 QIKKAGSICRPAEDCEDPENCTGSPACPKDQFRVNGFPCKNGBEGYCFMGKCPTRDQC 511
Db 452 QIKKAGSICRPAEDCEDPENCTGSPACPKDQFRVNGFPCKNGBEGYCFMGKCPTRDQC 511
Qy 517 TELWPGTEVAVKSWYRNEGSKYGYWRRVWVWTLIPKANWVWGCJLWMOGG--SWNLP 574
Db 517 TELWPGTEVAVKSWYRNEGSKYGYWRRVWVWTLIPKANWVWGCJLWMOGG--SWNLP 574
Qy 512 SELFDDDEATHSDICYKNTYKNGKFGYCKNKENRFLPCEEXDVRGKLIYCTGSELSYLL- 570
Db 512 SELFDDDEATHSDICYKNTYKNGKFGYCKNKENRFLPCEEXDVRGKLIYCTGSELSYLL- 570
Qy 575 WKGRIVTW-----LTWKTWBPWTSQFICGWANGTKWGNKKNWVINAENWVIEKAY 624
Db 575 WKGRIVTW-----LTWKTWBPWTSQFICGWANGTKWGNKKNWVINAENWVIEKAY 624
Qy 571 --GEDKTYHLKDPQONATVVKCTIFLVHDSITDGLVASGTGCGDGVNNGECLNMBKVY 628
Db 571 --GEDKTYHLKDPQONATVVKCTIFLVHDSITDGLVASGTGCGDGVNNGECLNMBKVY 628
Qy 625 KSTWSSKWKCHAVWHELQWQEBGTTPPWVWSSVWVHWSIVGVLPNPAVIVVIVAM 684
Db 625 KSTWSSKWKCHAVWHELQWQEBGTTPPWVWSSVWVHWSIVGVLPNPAVIVVIVAM 684
Qy 629 NSTNCPSQCHENPMDHGLQCHCEGQAPVAWEETLNTVAILIVLVLVIVIGVILIL 688
Db 629 NSTNCPSQCHENPMDHGLQCHCEGQAPVAWEETLNTVAILIVLVLVIVIGVILIL 688
Qy 685 VIRHOSREKOKWORPLSTTG-----TRP-----HKQRKXPQWKAQVP 724
Db 685 VIRHOSREKOKWORPLSTTG-----TRP-----HKQRKXPQWKAQVP 724
Qy 689 LIRYQKCIKLQVQSPPIETGVENKGVYFGDEQQWRTPEILPEHFLNQRTPESIESL-P 747
Db 689 LIRYQKCIKLQVQSPPIETGVENKGVYFGDEQQWRTPEILPEHFLNQRTPESIESL-P 747
Qy 725 QEMSCKPHEVYVLPVQGNPPAS 747
Db 725 QEMSCKPHEVYVLPVQGNPPAS 747
Qy 748 TSFS--SPHYITL-----KPAAS 762
Db 748 TSFS--SPHYITL-----KPAAS 762

RESULT 22
US-09-908-193-35
; Sequence 35, Application US/09908193
; Publication No. US20020192748A1
; GENERAL INFORMATION:
```

Db 505 TRNEQCSSELFAGAEESHSLSVRNNKGNRRFGYCKNKNFTFVPCBEKDKCKGKIYCSGR 564
Qy 570 -----SNLPMKGRIVTWLTWKTWPEWTSOEIGMVGANGTKWGNKVMWNAEWYMI 620
Db 565 PSSRLGENKTYNLKNVKNVT-1KCRTFMELHNSRDMGLVNSGTCKGDMVCSNGECIEH 623
Qy 621 EKAYKSTWNSKWKGHVWHELOHQBEGWTPPWVWSSVVVWHSIVVGVLPWMAVIMV 680
Db 624 EKAYNSTICSPCEDNDVDNEPECQCEGSIITWGEALNITSVIMVIL--VMVIGV 681
Qy 681 -VWAVISEQSRREKQKQRLSTTGT-----RPHKQKXZ 716
Db 682 GLVILLIRYQKCKMKQVSSPREIRGVENKGYFPEERQTRSEP 725

RESULT 13
US-09-921-823-17
; Sequence 17, Application US/09921823
; Patent No. US20020081685A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: DISINTEGRIN HOMOLOGS, ZSNK1C, ZSNK11,
; TITLE OF INVENTION: ZSNK12
; FILE REFERENCE: 03-55
; CURRENT APPLICATION NUMBER: US/09/921,823
; CURRENT FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 60/222,654
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 611
; TYPE: PRT
; ORGANISM: Sistrusus miliaris
US-09-921-823-17

Query Match 16.0%; Score 755.5; DB 9; Length 611;
Best Local Similarity 35.1%; Pred. No. 4.8e-52;
Matches 222; Conservative 104; Mismatches 276; Indels 31; Gaps 12;

Qy 1 MCGLLPVSLLSV---AVSAIKELPGVKYEVVYV-RLHPLHKEAKEPEQOEWE 56
Db 1 MCGVLL-VTCLAAFPYQSSILSGVNDVEVVTTRKVTALPKGA-----QPKIEDA 54
Qy 57 LKYNMTNGKIAVLKKNKMLLAPGYTETYNSTGKEITTSPOIMWVYQGHILNEKY 116
Db 55 MGYEFKMGEPVVLHKKNLFSKDYSETHSPDGROITTYPMIEDHCYHGRIONDAD 114
Qy 117 SWASISTWRLGYSQGWRYWIEPLSPIHGWQEHALWKYNPEKYNWSTWGMKGLW 176
Db 115 STASISACNGLGHFKLOGEMYLIPLKLPD--SEAHAVKYENITEKEDEAP-KMGVVT- 170
Qy 177 AHWLQONLALPATKLVKLRKVQ----BHEKYIEYLVLMNGEKRYNENOWEIRKRW 232
Db 171 QWESYE---PIKAFQNLTFEQAYLDKAYVEVWVLDHGMVYKYXDDLDKTRIV 227
Qy 233 EKANYVNMVLYKKNLTHVALVGMIEITWKKIKITPNASWLENWSKWRSVLSRRKHVI 292
Db 228 EIVWTMNEIYIPLNIRVALVHLEIWSNRDLINVSSAAGDTLGSFGEMWRETDLLRHKSHDN 287
Qy 293 AGLITATELAGTTVGLAWNSTWSP-YSVGVVQVHNSNLLRVRAGTVAHEHNGWGNW 351
Db 288 AGLLTTDFDGTGVLAY:SSMCQPSVGVIOBHSITNLLMAVTVVAHEHNGHNGSHDG 347
Qy 352 YSKWKPSTIIVWVKALSWIPTVWSSRLSYKWKWVKLSNWLNAPLPTWISTPIWG 411
Db 348 NQCHGAPSCIAERLSHOPSTQFSDCSEYCRITYLKNRRPQCILNEPLITDIVSPVCG 407
Qy 412 NQLVEMGEVWVWGTSENTINWAKTWIKIKATQWALGEMWKEKQWKKAGWVRDAKE 471
Db 408 NEELLEGEBCDCGSPANCPCDAATCKLTPGSQCAKGLCCQCRFKGKTECRAAKXD 467

Qy 472 WMLPEWNGKSGNWRVWVQVNGWPHHGHCKGHLMTWTPTLQEWTELWCPGTEVAKSW 531
Db 468 CDMADLCTGQSAKCTDRSQRNHGPLCLANKGYCYNRTCTPTMKNQCISFFGPSATVARDSC 527
Qy 532 YNRNKGSSYGVYVRRVWMTLIPWKNMTWGLWQGSWNLPWKGRIVTLWLTWKTWMP 591
Db 528 FXTNKGSSYGYCRKENGTKIPCEPDQYKGRLFCEYV--NKPCKKQNCNVI----YTP- 579
Qy 592 MTSQISGMVANGTKWGNKVMWNAEWVWIEKAY 624
Db 580 -TDEDIGMVLPTCKGRGKVCNNGHCVDTATAY 611
RESULT 14
US-10-232-972B-2
; Sequence 2, Application US/10232972B
; Publication No. US20040043387A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Hwan-Wup
; APPLICANT: Sun, Guang-Huan
; APPLICANT: Lin, Yu-Chi
; APPLICANT: Chang, Sun-Yuan
; TITLE OF INVENTION: Nucleic Acid Molecules and Polypeptides Related to h-ADAM 7
; FILE REFERENCE: 87161178-562001
; CURRENT APPLICATION NUMBER: US/10/232,972B
; CURRENT FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 775
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: LIPID
; LOCATION: {}..(775)
US-10-232-972B-2

Query Match 16.0%; Score 755; DB 12; Length 775;
Best Local Similarity 30.9%; Pred. No. 6.9e-52;
Matches 245; Conservative 136; Mismatches 335; Indels 76; Gaps 20;

Qy 1 MCGLLPVSLLSVAVSAIKE--LPGVKYEVVYVPIRLHPLHKEAKEPEQOE---QMET 55
Db 1 MFGCIEFMILL---IPQVKEFILLGVEGQOLVRPKKLPKIDGTGHTDDILKITYEE 57
Qy 56 ELKYMTNGKIAVLKKNKMLLAPGYTETYNSTGKEITTSPOIMWVYQGHILNEK 115
Db 58 ELLYEIKLRKTLVLHLRSREFLGSNYSSEYFYSKMGAPTRHPQIMDHCFYQGSIVHEY 117
Qy 116 VSWASISTWRLGYSQGWRYWIEPLSPIHGWQEHALWKYN--PWEKYNWSTWGMW 172
Db 118 DSAASISTCNGLRGPFPRINDQRYLIEPVK--YSGEHLVFKNLNRPVPGANYSCV--- 171
Qy 173 GVLWHLQONLALPATKLVKLRKVQEHKEKYIEYLVLMNGEKRYNENOWEIRKRW 232
Db 172 ---ELNFTKTPVGDNESEDSKIKGIHD-EKYVELFIVADDTVYRNGHPHKNLRIRW 227
Qy 233 EKANYVNMVLYKKNLTHVALVGMIEITWKKIKITPNASWLENWSKWRSVLSRRKHVI 292
Db 228 GNVNFMVMIYKTLNHTVLTGVIEIETHEDKIELYSNIETTLRFPSPQEKILATRKDFH 287
Qy 293 AGLITATELAGTTVGLAWNSTWSP-YSVGVVQVHNSNLLR---VAGTVAHEHNGWGN 347
Db 288 VLLSGKWLVSHVQGISYFGMCLPYISITIK----DLLDPTNIIANMAHQLGNLGM 343
Qy 348 MHWVWTSKWPSTIIVWVKALSWIPTVWSSRLSYKWKWVKLSNWLNAPLPTWIS 406
Db 344 QHDEPCTCFSGKCVWDSGS--IPALFKSKCSQNOYHVKDKYKPTCMLNIPFPYFHD 401
Qy 407 TPIWGNQVEMGEVWVWGTSENTINWAKTWIKIKATQWALGEMWKEKQWKKAGWVR 466
Db 402 FQFCGNKLLDEGEBCDCGPAQECTNCCDAHTCVLKPGETCAEGSCCESQIKKAGSICR 461

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2004, 17:50:35 ; Search time 60 Seconds
(without alignments)
3649.572 Million cell updates/sec

Title: US-10-726-148A-15
Perfect score: 4712
Sequence: 1 LQGLPVSLLSVAVSAIK.....PPTWKWNPMTPKWSNPKA 775

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 29Jan04: *
1: Geneseqp1980s: *
2: Geneseqp1990s: *
3: Geneseqp2000s: *
4: Geneseqp2002s: *
5: Geneseqp2003as: *
6: Geneseqp2003bs: *
7: Geneseqp2003bs: *
8: Geneseqp2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3128	66.4	775	3 AAW90855	Aaw90855 Human ADA
2	3128	66.4	788	4 ABB11329	Abb11329 Human eMD
3	3123	66.3	775	3 AAW90865	Aaw90865 Human ADA
4	2132	45.2	540	3 AAW90851	Aaw90851 Human ADA
5	2132	45.2	540	3 AAW90864	Aaw90864 Human ADA
6	2098	44.5	530	2 AAW14772	Aaw14772 Human met
7	1610	34.2	439	4 AAG73920	Aag73920 Human col
8	1605	34.1	435	4 AAB50937	Aab50937 ADAM prot
9	1605	34.1	435	4 AAB64681	Aab64681 Human sec
10	1605	34.1	435	6 ABB06027	Abb06027 Human nov
11	1284	27.2	367	4 AAB64745	Aab64745 Human sec
12	854	18.1	233	3 AAB53474	Aab53474 Human col
13	817	17.3	201	3 AAW90853	Aaw90853 Human ADA
14	815	17.3	208	4 AAB50938	Aab50938 ADAM prot
15	815	17.3	208	6 ABB06028	Abb06028 Human nov
16	759.5	16.1	776	4 AAB47567	Aab47567 Protease
17	755.5	16.0	611	5 AAU76794	Aau76794 Pigmy rat
18	755.5	16.0	789	7 ADE63953	Ade63953 Rat Prote
19	750.5	15.9	779	5 AAU77409	Aau77409 Human NOV
20	750	15.9	754	7 ADE63955	Ade63955 Human Pro
21	745	15.8	778	5 AAU77408	Aau77408 Human NOV
22	736	15.6	695	4 AAU16950	Aau16950 Human nov
23	735.5	15.6	613	2 AAW73009	Aaw73009 Cobra ven
24	732.5	15.5	621	2 AAW73013	Aaw73013 Cobra ven
25	717	15.2	621	2 AAW73007	Aaw73007 Cobra ven

ALIGNMENTS

RESULT 1
AAW90855
ID AAW90855 standard; protein; 775 AA.

XX AAW90855;
DT 28-JUN-2000 (first entry)
XX Human ADAM protein #2.
XX ADAM protein; human; A disintegrin and metalloprotease; diagnosis;
KW drug development; intervertebral hernia; sciatica; pulmonary fibrosis;
KW diabetic nephropathy; hepatic fibrosis; glomerulitis; osteopetrosis.
XX Homo sapiens.

XX WO200014227-A1.
PD 16-MAR-2000.
XX 02-SEP-1993; 99WO-JP004766.
XX 03-SEP-1998; 98JP-00250115.
XX (TAKE) TAKEDA CHEM IND LTD.

XX Yoshimura K, Hikichi Y, Nishimura A;
WPI: 2000-271056/23.
N-PSDB; AA289683.

Novel protein belong to A disintegrin and metalloprotease family, with
protease activity and extracellular matrix digesting enzyme activity, for
gene diagnosis and developing drugs for treating e.g. sciatica and
glomerulitis.

Claim 4; Page 98-102; 109pp; Japanese.

This invention describes a novel human protein (I) which belongs to the
ADAM (A disintegrin and metalloprotease) protein family. The protein,
peptide fragment and antibody are useful for gene diagnosis and in the
development of drugs to prevent or treat intervertebral hernia, sciatica,
glomerulitis, diabetic nephropathy, hepatic fibrosis, pulmonary fibrosis
or osteopetrosis. This sequence represents a human ADAM protein which is
described in the method of the invention
Sequence 775 AA;

Aaw28509 Product O
Adc38650 Human sec
Aaw75769 Human met
Aaw73011 Cobra ven
Ada27208 Murine pr
Aaw73010 Cobra ven
Ada27196 Human MD8
Aab20251 Human ADA
Aaw86156 Human PRO
Aad78889 Human PRO
Add18580 Human dis
Abu61896 E. carina
Aaw80066 Saw-scale
Aaw73008 Cobra ven
Aay79413 Southern
Aae03036 Agkistrod
Aau76795 Cottonmou
Aab50935 ADAM prot
Aae19181 Human pro
Aae24351 Human nov

Query Match	66.4%;	Score 3128;	DB 3;	Length 775;
Best Local Similarity	84.4%;	Pred. No. 1.2e-245;		
Matches 654;	Conservative 29;	Mismatches 99;	Indels 0;	Gaps 0;
QY 1	MLQGLLPVSLLSVAVSAIKELPGVKYEVYVPIRLHPLHREAKEPQEQWETELKYK 60			
DB 1	MLQGLLPVSLLSVAVSAIKELPGVKYEVYVPIRLHPLHREAKEPQEQWETELKYK 60			
QY 61	MTINGKIAVLKKNKLLAGYTYTYSNKGKETTSPQWMMWYOGHILNEKVSAS 120			
DB 61	MTINGKIAVLKKNKLLAGYTYTYSNKGKETTSPQWMMWYOGHILNEKVSAS 120			
QY 121	ISTWGLRGYWSQGRYWIPLSPHRSWQGEHALWKYNPWEKYNWSTWGMVGLWAHL 180			
DB 121	ISTWGLRGYWSQGRYWIPLSPHRSWQGEHALWKYNPWEKYNWSTWGMVGLWAHL 180			
QY 181	QONIALPATKLVKLRKQVQEHKEIYIYVLWNGEWKRYNENOWEIRKRVWEMANYNM 240			
DB 181	QONIALPATKLVKLRKQVQEHKEIYIYVLWNGEWKRYNENOWEIRKRVWEMANYNM 240			
QY 241	LYKLNTHVALVGEIWTWKIKITPNASVTLNKKWRSVLSRRKRHWLAQILITATE 300			
DB 241	LYKLNTHVALVGEIWTWKIKITPNASVTLNKKWRSVLSRRKRHWLAQILITATE 300			
QY 301	LAGITVGLAFNMTWMSPIYSVGVVQHSNLLRVAGTMAHEGHNWGMWYSWKWPSTI 360			
DB 301	LAGITVGLAFNMTWMSPIYSVGVVQHSNLLRVAGTMAHEGHNWGMWYSWKWPSTI 360			
QY 361	WMMKALSNIYPTWMSNRSLSYKRWKWSKLSNWLWNPAPLPTWIIISTPIGNQLVEMGEM 420			
DB 361	WMMKALSNIYPTWMSNRSLSYKRWKWSKLSNWLWNPAPLPTWIIISTPIGNQLVEMGEM 420			
QY 421	WMMKALSNIYPTWMSNRSLSYKRWKWSKLSNWLWNPAPLPTWIIISTPIGNQLVEMGEM 420			
DB 421	WMMKALSNIYPTWMSNRSLSYKRWKWSKLSNWLWNPAPLPTWIIISTPIGNQLVEMGEM 420			
QY 481	KSGNRPWRRVQNGPWHHKGELWMTPTLQEQWTELWGPSTEVAKSWTNRRNEGSK 540			
DB 481	KSGNRPWRRVQNGPWHHKGELWMTPTLQEQWTELWGPSTEVAKSWTNRRNEGSK 540			
QY 541	YGWRRVWTLIPKANTMTGKLWQGSNLPKGRIVTWLTKTWPEWTSOIGMVK 600			
DB 541	YGWRRVWTLIPKANTMTGKLWQGSNLPKGRIVTWLTKTWPEWTSOIGMVK 600			
QY 601	ANGTKGKNGKYNIAEWVIEKAYKSTWSSKWKGHVAVNHQWEGEWTTPWMMWSS 660			
DB 601	ANGTKGKNGKYNIAEWVIEKAYKSTWSSKWKGHVAVNHQWEGEWTTPWMMWSS 660			
QY 661	VVWMSIYVGVLMPMVAVIWWVAVIRHQSREKQKQRPSTTTGTTPHKKRPQVVK 720			
DB 661	VVWMSIYVGVLMPMVAVIWWVAVIRHQSREKQKQRPSTTTGTTPHKKRPQVVK 720			
QY 721	AVQOQMSQMKHVTWLVPEGNEPPASHKWTNALPPTVWKNPNMSTPKNSPKA 775			
DB 721	AVQOQMSQMKHVTWLVPEGNEPPASHKWTNALPPTVWKNPNMSTPKNSPKA 775			
RESULT 2				
AB111929				
ID	AB111929	standard;	peptide; 788 AA.	
XX	AB111929			
AC	AB111929			
XX	AB111929			
DT	11-JAN-2002	(first entry)		
XX	Human eMDC II protein homologue, SEQ ID NO:2299.			
XX	Human; cytokine; cell proliferation; cell differentiation; growth factor;			
XX	haematopoiesis regulation; tissue growth; immunomodulator; activin;			
XX	inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;			
XX	proliferation; metastasis; cancer; tumour; haematopoietic disorder;			
XX	myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;			
XX	chronic inflammatory condition; proliferative retinopathy;			

Query Match 66.4%; Score 3128; DB 4; Length 788;

Best Local Similarity 84.4%; Pred. No. 1.3e-245; Matches 654; Conservative 29; Mismatches 92; Indels 0; Gaps 0;	
QY	1 MLOGLPVSLLSVAVSAIKELPGVKYEVVYPIRLHPLHREAKEPEQOQOMETELKYK 60
DB	14 MLOGLPVSLLSVAVSAIKELPGVKYEVVYPIRLHPLHREAKEPEQOQOFETELKYK 73
QY	61 MTINGKIAVLVYLNKKNLLAPGYTETYNSTGKEITTSPOIMDDCYQGHILNEKVSWS 120
DB	74 MTINGKIAVLVYLNKKNLLAPGYTETYNSTGKEITTSPOIMDDCYQGHILNEKVSWS 133
QY	121 ISTWGLRGVWSQWORYWIEPLSPHRWGQEHALWKNYNPWKNYSTWGMGVLWAHNL 180
DB	134 ISTCGLRGYFQGDQRYFIEPLSPHRDQGEHALFKYNPDEKNYSTCGMDGVLWAHDL 193
QY	181 QONIALPATKLVKLRKVKQEHKEIYEVYLVLMNGEKKRYNENQWIEIRKRVEMANYVM 240
DB	194 QONIALPATKLVKLRKVKQEHKEIYEVYLVLMNGEKKRYNENQWIEIRKRVEMANYVM 253
QY	241 LYKKNLTHVALVGMIEIWTWKIKITPNASFTLENFSKWRGSLSRKRRHWAQILITATE 300
DB	254 LYKKNLTHVALVGMIEIWTWKIKITPNASFTLENFSKWRGSLSRKRRHWAQILITATE 313
QY	301 LAGTTVGLAWMTMSPYSVGVVQVHNSWLLRVAGTMAHEMGNWGMWYKWKWPSTII 360
DB	314 LAGTTVGLAFMTMSPYSVGVVQVHNSWLLRVAGTMAHEMGNWGMWYKWKWPSTII 373
QY	361 WYMWKALSFIPTWSSWSRLSYKWKVWKLNLNAPLPTWIIISTPWGNQVEMGEM 420
DB	374 CYMDKALSFIPTDFSSCSRLSYKWKVWKLNLNAPLPTWIIISTPWGNQVEMGEM 433
QY	421 WYMWKALSFIPTWSSWSRLSYKWKVWKLNLNAPLPTWIIISTPWGNQVEMGEM 480
DB	434 CYMDKALSFIPTDFSSCSRLSYKWKVWKLNLNAPLPTWIIISTPWGNQVEMGEM 493
QY	481 KSGNPPWPKVQVNGWPKHGHGHLMGITWFTLQEOHTELWGPTEVAMKSWYNRNEGSK 540
DB	494 KSGNPPWPKVQVNGWPKHGHGHLMGITWFTLQEOHTELWGPTEVAMKSWYNRNEGSK 553
QY	541 YGYRRVWVWTLIPKANKWTKMGLWQSGSNLWPKGRIVTWLTKTWKWPETSOBGMV 600
DB	554 YGYRRVWVWTLIPKANKWTKMGLWQSGSNLWPKGRIVTWLTKTWKWPETSOBGMV 613
QY	601 ANGTWGNKKNVINAEMVWIEKAYKSTNNSKWKCHAVWHELOWGHEGTPPWVWNS 660
DB	614 ANGTWGNKKNVINAEMVWIEKAYKSTNNSKWKCHAVWHELOWGHEGTPPWVWNS 673
QY	661 VYVHWSIVVGVLPMAVIVVWVAVIRHQSREKQKQWRLPSTCTTRPHKQKRPQMYK 720
DB	674 VYVHWSIVVGVLPMAVIVVWVAVIRHQSREKQKQWRLPSTCTTRPHKQKRPQMYK 733
QY	721 AVQOEMSQMKHVVYLPVEGNEPPASWHTNAPLPTWKNVMPSTPKWSNPKA 775
DB	734 AVQOEMSQMKHVVYLPVEGNEPPASWHTNAPLPTWKNVMPSTPKWSNPKA 788
RESULT 3	
ID	AAW90865
XX	AAW90865 standard; protein; 775 AA.
AC	AAW90865;
XX	AAW90865;
DT	28-JUN-2000 (first entry).
DE	Human ADAM protein #4.
KW	ADAM protein; human; A disintegrin and metalloprotease; diagnosis;
KW	drug development; intervertebral hernia; sciatica; pulmonary fibrosis;
KW	diabetic nephropathy; hepatic fibrosis; glomerulitis; osteopetrosis.
OS	Homo sapiens.
XX	XX
PN	WO200014227-A1.
Query Match 66.3%; Score 3123; DB 3; Length 775; Best Local Similarity 84.3%; Pred. No. 3.2e-245; Matches 653; Conservative 29; Mismatches 93; Indels 0; Gaps 0;	
QY	1 MLOGLPVSLLSVAVSAIKELPGVKYEVVYPIRLHPLHREAKEPEQOQOMETELKYK 60
DB	1 MLOGLPVSLLSVAVSAIKELPGVKYEVVYPIRLHPLHREAKEPEQOQOFETELKYK 60
QY	61 MTINGKIAVLVYLNKKNLLAPGYTETYNSTGKEITTSPOIMDDCYQGHILNEKVSWS 120
DB	61 MTINGKIAVLVYLNKKNLLAPGYTETYNSTGKEITTSPOIMDDCYQGHILNEKVSWS 120
QY	121 ISTWGLRGVWSQWORYWIEPLSPHRWGQEHALWKNYNPWKNYSTWGMGVLWAHNL 180
DB	121 ISTCGLRGYFQGDQRYFIEPLSPHRDQGEHALFKYNPDEKNYSTCGMDGVLWAHDL 180
QY	181 QONIALPATKLVKLRKVKQEHKEIYEVYLVLMNGEKKRYNENQWIEIRKRVEMANYVM 240
DB	181 QONIALPATKLVKLRKVKQEHKEIYEVYLVLMNGEKKRYNENQWIEIRKRVEMANYVM 240
QY	241 LYKKNLTHVALVGMIEIWTWKIKITPNASFTLENFSKWRGSLSRKRRHWAQILITATE 300
DB	241 LYKKNLTHVALVGMIEIWTWKIKITPNASFTLENFSKWRGSLSRKRRHWAQILITATE 300
QY	301 LAGTTVGLAWMTMSPYSVGVVQVHNSWLLRVAGTMAHEMGNWGMWYKWKWPSTII 360
DB	301 LAGTTVGLAWMTMSPYSVGVVQVHNSWLLRVAGTMAHEMGNWGMWYKWKWPSTII 360
QY	361 WYMWKALSFIPTWSSWSRLSYKWKVWKLNLNAPLPTWIIISTPWGNQVEMGEM 420
DB	361 CYMDKALSFIPTDFSSCSRLSYKWKVWKLNLNAPLPTWIIISTPWGNQVEMGEM 420
QY	421 WYMWKALSFIPTWSSWSRLSYKWKVWKLNLNAPLPTWIIISTPWGNQVEMGEM 480
DB	421 CYMDKALSFIPTDFSSCSRLSYKWKVWKLNLNAPLPTWIIISTPWGNQVEMGEM 480
QY	481 KSGNPPWPKVQVNGWPKHGHGHLMGITWFTLQEOHTELWGPTEVAMKSWYNRNEGSK 540
DB	481 KSGNPPWPKVQVNGWPKHGHGHLMGITWFTLQEOHTELWGPTEVAMKSWYNRNEGSK 540
QY	541 YGYRRVWVWTLIPKANKWTKMGLWQSGSNLWPKGRIVTWLTKTWKWPETSOBGMV 600

D6 541 YGCRVDDTLIPCKANDTMCKGLPCDGGDNLPHKGRIVTFLTKTTPDPTSOBICMV 600
 QY 601 ANGTWGNKWKWIKAEVWYIEKAYKSTNWSSKWKGHAVWVHQLQWQWEGNTPPPWWS 660
 DE 601 ANGTGCGDNKVCINAEVCDIEKAYKSTNCSSKCKGHAVCDHELQCCCEGWIPPCDDSS 660
 QY 661 VVWHSIVVGLWPKAVIVVWVAVIRHQSSEKOKKQRELSITGTRPHKOKRKPQWVK 720
 D6 661 VVHFHSIVVGLWPKAVIVVWVAVIRHQSSEKOKKQRELSITGTRPHKOKRKPQWVK 720
 QY 721 AVQPEMSQMKPHVWVWLPVEGNEPPASWHEKTNALPPTVWKNPMTSPKWSNPKA 775
 D6 721 AVQPEMSQMKPHVWVWLPVEGNEPPASWHEKTNALPPTVWKNPMTSPKWSNPKA 775

RESULT 4

AAW90851
 ID AAW90851 standard; protein; 540 AA.

AC AAW90851;

DT 28-JUN-2000 (first entry)

DE Human ADAM protein #1.

KW ADAM protein; human; A disintegrin and metalloprotease; diagnosis;
 KW drug development; intervertebral hernia; sciatica; pulmonary fibrosis;
 KW diabetic nephropathy; hepatic fibrosis; glomerulitis; osteopetrosis.

OS Homo sapiens.

PN WO200014227-A1.

PD 16-MAR-2000.

PF 02-SEP-1999; 99WO-JP004766.

PR 03-SEP-1998; 98JP-00250115.

FA (TAKE) TAKEDA CHEM IND LTD.

PI Yoshimura K, Hikichi Y, Nishimura A;

XX WPI; 2000-271056/23.

PT Novel protein belong to A disintegrin and metalloprotease family, with
 PT protease activity and extracellular matrix digesting enzyme activity, for
 PT gene diagnosis and developing drugs for treating e.g. sciatica and
 PT glomerulitis.

PS Claim 4; Page 90-92; 109pp; Japanese.

CC This invention describes a novel human protein (I) which belongs to the
 CC ADAM (A disintegrin and metalloprotease) protein family. The protein,
 CC peptide fragment and antibody are useful for gene diagnosis and in the
 CC development of drugs to prevent or treat intervertebral hernia, sciatica,
 CC glomerulitis, diabetic nephropathy, hepatic fibrosis, pulmonary fibrosis
 CC or osteopetrosis. This sequence represents a human ADAM protein which is
 CC described in the method of the invention

SQ Sequence 540 AA;

Query Match 45.2%; Score 2132; DB 3; Length 540;

Best Local Similarity 84.9%; Pred. No. 8.5e-165;

Matches 444; Conservative 20; Mismatches 59; Indels 0; Gaps 0;

QY 1 MQGLLPVSLLSVAVSAIKELPGVKYEVVYPIRLHLHKEAKEPEQOQOWETELKYK 60

D6 1 MQGLLPVSLLSVAVSAIKELPGVKYEVVYPIRLHLHKEAKEPEQOQOWETELKYK 60

QY 61 MTINGKIAVLKKNKKNLLAPGYTETTYNNTSGKEITTSQIMWVYQGHILNEKYSWAS 120

D6 61 MTINGKIAVLKKNKKNLLAPGYTETTYNNTSGKEITTSQIMWVYQGHILNEKYSWAS 120

QY 121 STWGLRGYWSQGWRYWIEPLSPIHRWQOHALWKYNPNKKNYVSTWGMGCVLWAHKL 180
 D6 121 STWGLRGYWSQGWRYWIEPLSPIHRWQOHALWKYNPNKKNYVSTWGMGCVLWAHKL 180
 QY 181 QONIALPATKLVKLRKVRQVEHEKYEYVVLWNGEWKRYNENOMERKRVWEMANYVM 240
 D6 181 QONIALPATKLVKLRKVRQVEHEKYEYVVLWNGEWKRYNENOMERKRVWEMANYVM 240
 QY 241 LYXKLNTHVALVGBEIBWTKWIKITPNASWTLNWSKWRGSLSRKRRHWIAQLITATE 300
 D6 241 LYXKLNTHVALVGBEIBWTKWIKITPNASWTLNWSKWRGSLSRKRRHWIAQLITATE 300
 QY 301 LAGTTVGLAKWMTMSPYSVGVVQVHNSNLLVAGTMAHEMGNHGNWMTSWKWPSTI 360
 D6 301 LAGTTVGLAKWMTMSPYSVGVVQVHNSNLLVAGTMAHEMGNHGNWMTSWKWPSTI 360
 QY 361 VYMKKALSMYIPTWSSWSRLSYWKWVKWLNWLNADLPPTWIIISTPIWGNQVEMGEW 420
 D6 361 VYMKKALSMYIPTWSSWSRLSYWKWVKWLNWLNADLPPTWIIISTPIWGNQVEMGEW 420
 QY 421 WWTGTSEWTNIWNAKTKIKATQWALGEWKEWKKAGVWVRPAKWVWLPDMWNG 480
 D6 421 WWTGTSEWTNIWNAKTKIKATQWALGEWKEWKKAGVWVRPAKWVWLPDMWNG 480
 QY 481 KSGNFWVRWQVWHPWHGKHWTLMGTWPTLOEOWTELWGP 523
 D6 481 KSGNFWVRWQVWHPWHGKHWTLMGTWPTLOEOWTELWGP 523
 RESULT 5
 AAW90864
 ID AAW90864 standard; protein; 540 AA.
 AC AAW90864;
 DT 28-JUN-2000 (first entry)
 DE Human ADAM protein #3.
 KW ADAM protein; human; A disintegrin and metalloprotease; diagnosis;
 KW drug development; intervertebral hernia; sciatica; pulmonary fibrosis;
 KW diabetic nephropathy; hepatic fibrosis; glomerulitis; osteopetrosis.
 OS Homo sapiens.
 PN WO200014227-A1.
 PD 16-MAR-2000.
 PF 02-SEP-1999; 99WO-JP004766.
 PR 03-SEP-1998; 98JP-00250115.
 FA (TAKE) TAKEDA CHEM IND LTD.
 PI Yoshimura K, Hikichi Y, Nishimura A;
 XX WPI; 2000-271056/23.
 PT Novel protein belong to A disintegrin and metalloprotease family, with
 PT protease activity and extracellular matrix digesting enzyme activity, for
 PT gene diagnosis and developing drugs for treating e.g. sciatica and
 PT glomerulitis.
 PS Claim 4; Page 90-92; 109pp; Japanese.
 CC This invention describes a novel human protein (I) which belongs to the
 CC ADAM (A disintegrin and metalloprotease) protein family. The protein,
 CC peptide fragment and antibody are useful for gene diagnosis and in the
 CC development of drugs to prevent or treat intervertebral hernia, sciatica,
 CC glomerulitis, diabetic nephropathy, hepatic fibrosis, pulmonary fibrosis
 CC or osteopetrosis. This sequence represents a human ADAM protein which is
 CC described in the method of the invention

CC or osteopetrosis. This sequence represents a human ADAM protein which is
 CC described in the method of the invention
 XX SQ Sequence 540 AA;

Query Match 45.2%; Score 2132; DB 3; Length 540;
 Best Local Similarity 84.9%; Pred. No. 8.5e-165;
 Matches 444; Conservative 20; Mismatches 59; Indels 0; Gaps 0;

QY 1 MLOGLPVSLLSVAISAIAKELPGVKYEVVYPIRLPLHPLHREAKEPEQOEQWETELKYK 60
 DB 1 MLOGLPVSLLSVAISAIAKELPGVKYEVVYPIRLPLHPLHREAKEPEQOEQWETELKYK 60
 QY 61 MTINGKIAVLKKNKLLAPGYTETYNSTGKEITTSPOIMMWYQGHILNEKVSWS 120
 DB 61 MTINGKIAVLKKNKLLAPGYTETYNSTGKEITTSPOIMMWYQGHILNEKVSWS 120
 QY 121 ISTWRGLRGYVSGQGRYIEPLSPIHRWGQEHALMKYNPWEKNYVSTWGMGVLMAHVL 180
 DB 121 ISTWRGLRGYVSGQGRYIEPLSPIHRWGQEHALMKYNPWEKNYVSTWGMGVLMAHVL 180
 QY 181 QONIALPATKLVKLRKQVQEHKEIYEVYLVWNGEWMKRYNENQWEIRKRVEMANYVM 240
 DB 181 QONIALPATKLVKLRKQVQEHKEIYEVYLVWNGEWMKRYNENQWEIRKRVEMANYVM 240
 QY 241 LYKLNTHVALVGMETWTKWKIKITPNASWTLENWSKRGSVLSRRKRHWIAQLITATE 300
 DB 241 LYKLNTHVALVGMETWTKWKIKITPNASWTLENWSKRGSVLSRRKRHWIAQLITATE 300
 QY 301 LAGTTVGLAWMTMSPYSVGVMQHSNLLRVAGTMAHEMGNHGMWYVSTWGMGVLMAHVL 360
 DB 301 LAGTTVGLAWMTMSPYSVGVMQHSNLLRVAGTMAHEMGNHGMWYVSTWGMGVLMAHVL 360
 QY 361 WYWKALSYIPTWSSWSRLSYKWKWKLNLNAPLPTWIIISTPIGNQVLEMGW 420
 DB 361 WYWKALSYIPTWSSWSRLSYKWKWKLNLNAPLPTWIIISTPIGNQVLEMGW 420
 QY 421 WNGTSEWTNIWNAKTWKIKATQWALGEMWKEKQCFKAGMVCPRPAKDECDLPENCG 480
 DB 421 WNGTSEWTNIWNAKTWKIKATQWALGEMWKEKQCFKAGMVCPRPAKDECDLPENCG 480
 QY 481 KSGNPPWVRWQVNGPWHHKGHLWMTPTLOEQWTELWGP 523
 DB 481 KSGNPPWVRWQVNGPWHHKGHLWMTPTLOEQWTELWGP 523

RESULT 6

AAW14772
 ID AAW14772 standard; protein; 530 AA.
 AC AAW14772;
 DT 19-JUN-1997 (first entry)
 DE Human metalloproteinase.
 KW Metalloproteinase; infectious disease; inflammation; autoimmune disease;
 XX allergy; atopic disease; cancer.
 OS Homo sapiens.
 XX Key Location/Qualifiers
 FT Domain 199..399
 FT /label= Catalytic_domain
 XX W09709430-A2.
 XX 13-MAR-1997.
 XX 05-SEP-1996; 96W0-GB002181.
 XX 05-SEP-1995; 95GB-00018023.
 XX 20-OCT-1995; 95GB-00021495.

PR 20-OCT-1995; 95GB-00021498.
 PR 21-DEC-1995; 95GB-00026229.
 PR 11-JUN-1996; 96GB-00012150.
 XX (CLT) CELLTECH THERAPEUTICS LTD.
 XX DoCherty AJP, Slocombe PM;
 XX WPI; 1997-192905/17.
 XX N-PSDB; AAT63123, AAT63124.
 XX New isolated human metalloproteinase - used to develop prods. for the
 PT treatment of e.g. infectious disease, inflammatory disease, autoimmunity
 PT or allergic, atopic diseases.
 XX Claim 4; Page 6; 16pp; English.
 XX A metalloproteinase (AAW14772) is the product of an isolated human DNA
 CC (AAT63123). The enzyme, or its catalytic domain, can be produced in
 CC prokaryotic or eukaryotic host cells and used to screen for cpds. that
 CC regulate its activity. Such cpds. can be used to treat diseases or
 CC disorders in human in which the over- or unregulated prodn. of
 CC metalloproteinase is implicated. The metalloproteinase can also be used
 CC to raise antibodies capable of regulating the activity of the enzyme
 XX SQ Sequence 530 AA;

Query Match 44.5%; Score 2098; DB 2; Length 530;
 Best Local Similarity 83.7%; Pred. No. 4.9e-162;
 Matches 438; Conservative 20; Mismatches 55; Indels 10; Gaps 1;

QY 1 MLOGLPVSLLSVAISAIAKELPGVKYEVVYPIRLPLHPLHREAKEPEQOEQWETELKYK 60
 DB 1 MLOGLPVSLLSVAISAIAKELPGVKYEVVYPIRLPLHPLHREAKEPEQOEQWETELKYK 60
 QY 61 MTINGKIAVLKKNKLLAPGYTETYNSTGKEITTSPOIMMWYQGHILNEKVSWS 120
 DB 61 MTINGKIAVLKKNKLLAPGYTETYNSTGKEITTSPOIMMWYQGHILNEKVSWS 120
 QY 121 ISTWRGLRGYVSGQGRYIEPLSPIHRWGQEHALMKYNPWEKNYVSTWGMGVLMAHVL 180
 DB 121 ISTWRGLRGYVSGQGRYIEPLSPIHRWGQEHALMKYNPWEKNYVSTWGMGVLMAHVL 180
 QY 181 QONIALPATKLVKLRKQVQEHKEIYEVYLVWNGEWMKRYNENQWEIRKRVEMANYVM 240
 DB 181 QONIALPATKLVKLRKQVQEHKEIYEVYLVWNGEWMKRYNENQWEIRKRVEMANYVM 240
 QY 241 LYKLNTHVALVGMETWTKWKIKITPNASWTLENWSKRGSVLSRRKRHWIAQLITATE 300
 DB 241 LYKLNTHVALVGMETWTKWKIKITPNASWTLENWSKRGSVLSRRKRHWIAQLITATE 300
 QY 301 LAGTTVGLAWMTMSPYSVGVMQHSNLLRVAGTMAHEMGNHGMWYVSTWGMGVLMAHVL 360
 DB 301 LAGTTVGLAWMTMSPYSVGVMQHSNLLRVAGTMAHEMGNHGMWYVSTWGMGVLMAHVL 360
 QY 361 WYWKALSYIPTWSSWSRLSYKWKWKLNLNAPLPTWIIISTPIGNQVLEMGW 420
 DB 361 WYWKALSYIPTWSSWSRLSYKWKWKLNLNAPLPTWIIISTPIGNQVLEMGW 420
 QY 421 WNGTSEWTNIWNAKTWKIKATQWALGEMWKEKQCFKAGMVCPRPAKDECDLPENCG 480
 DB 421 WNGTSEWTNIWNAKTWKIKATQWALGEMWKEKQCFKAGMVCPRPAKDECDLPENCG 480
 QY 481 KSGNPPWVRWQVNGPWHHKGHLWMTPTLOEQWTELWGP 523
 DB 481 KSGNPPWVRWQVNGPWHHKGHLWMTPTLOEQWTELWGP 523
 XX RESULT 7
 XX AAG73920
 XX ID AAG73920 standard; protein; 439 AA.
 XX AAG73920;
 XX AC AAG73920;

```
XX 03-SEP-2001 (first entry)
DC Human colon cancer antigen protein SEQ ID NO:4684.
DE Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
KW Homo sapiens.
OS WO200122320-A2.
PN 05-APR-2001.
XX 28-SEP-2000; 2000WO-US22524.
XX 29-SEP-1999; 99US-0157:37P.
PR 03-NOV-1999; 99US-0163280P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI WPI; 2001-235357/24.
DR N-PSDB; AAH33351.
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.
XX Claim 11; Page 6485-6487; 9803pp; English.
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patient's own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 692 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922
XX Sequence 439 AA;
XX Query Match 34.2%; Score 1610; DB 4; Length 439;
XX Best Local Similarity 79.1%; Pred. No. 1.9e-122;
XX Matches 345; Conservative 20; Mismatches 71; Indels 0; Gaps 0;
QY 340 EMGNWGMWHTWYKWKWPSSTVWVKALSYVPTWWSWSRLSYKWKWKLNLWLNAP 399
DB 4 EMGNFMGHHDDYKCKPCSTVMDKALSFYPTDFSSCSRLSYDKFEDKLSNCLFNAP 63
QY 400 LPTWIIISTPIWGNQVEMGEMWMMGTSEETWNIWWAKTWKIKATWQWALGEMWKKQWK 459
DB 64 LPTDIIISTPICGNQLVEMGEDCDGTSSECTNICCDAKTCKIKATFCALGECCKQFK 123
QY 460 KAGWVPAKWEWLPENWNGKSNWPNWQVNGWFWHKGKGLWLMGTWPTLQEQWTEL 519
DB 124 KAGWVCPAPKDECDLPENWNGKSNWPNWQVNGWFWHKGKGLWLMGTWPTLQEQWTEL 183
QY 520 WGPCTEVANKSWYNNRNGSSKYGYWRRVWWTLPKANKWNTWGLWQGGSNLPWKGR 579
DB 184 WGPCTEVADKSCYNNRNGSSKYGYCRRVDDTLIPKANDTMCGLFCQGGSDNLPWKGR 243
QY 580 VTWLTWKTWPEWTQSIEGMVANGTKWGNKVINAEWVWIEKAYKSTNWSKWKSHAV 639
DB 244 VTELTKTTPEDTSQIEGMVANGTKGDNKVCINAECDVIEKAYKSTNWSKWKSHAVC 303
QY 640 WHELQWQWEGWTPPPWWSVWVWHSIVVGVLPMAVIVVWVWVIRHOSRSREKQKQW 699
DB 304 DHELQWQWEGWTPPPWWSVWVWHSIVVGVLPMAVIVVWVWVIRHOSRSREKQKQW 363
QY 700 RPLSTTGTTRPHKOKRKPQWKAQVQPMQSKMHPVYMLPVEGNEPPASFWHKTNALPPTV 759
DB 364 RPLSTTGTTRPHKOKRKPQWKAQVQPMQSKMHPVYMLPVEGNEPPASFWHKTNALPPTV 423
QY 760 WKWNPMTSPKWSNPKA 775
DB 424 FKDNPMSTPKDSNPKA 439
RESULT 8
AAB50937
ID AAB50937 standard; protein; 435 AA.
XX AAB50937;
XX 20-MAR-2001 (first entry)
XX ADAM protein #3.
DE ADAM; disintegrin; metalloprotease; food additive; breast cancer;
KW ovarian.
XX Homo sapiens.
XX WO200073323-A2.
XX 07-DEC-2000.
XX 25-MAY-2000; 2000WO-US014308.
XX 27-MAY-1999; 99US-0136388P.
PR 09-JUL-1999; 99US-0142930P.
PR 28-JAN-2000; 2000US-0178717P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Wei P, Ni J, Hastings GA, Shi Y;
PI WPI; 2001-016507/02.
XX Seven nucleic acid molecules encoding ADAM polypeptides containing a
XX disintegrin and metalloprotease domain, useful in the prevention,
XX treatment and diagnosis of cancer, immune disorders, cardiovascular
XX disorders and neurological diseases.
XX Claim 11; Page 276-277; 287pp; English.
XX The present invention relates to seven members of the ADAMs (proteins
XX which contain A Disintegrin And Metalloprotease domain) protein family.
XX The ADAMs proteins and DNA may be used to treat disease, as a food
XX additive or preservative, for chromosome identification, as probes for
XX diagnosing a disorder related to the female reproductive system,
XX particularly breast and/or ovary cancer. They are also useful in the gene
XX therapy of breast and ovarian cancer
XX Sequence 435 AA;
XX Query Match 34.1%; Score 1605; DB 4; Length 435;
XX Best Local Similarity 79.1%; Pred. No. 4.9e-122;
XX Matches 344; Conservative 20; Mismatches 71; Indels 0; Gaps 0;
QY 341 MGNHMGWHHWYKWKWPSSTVWVKALSWYPTWWSWSRLSYKWKWKLNLWLNAPL 400
DB 1 MGNFMGHHDDYKCKPCSTVMDKALSFYPTDFSSCSRLSYDKFEDKLSNCLFNAPL 60
QY 401 PTWIIISTPIWGNQVEMGEMWMMGTSEETWNIWWAKTWKIKATWQWALGEMWKKQWK 460
```

Db 61 PTDIISTPIGQNLVEMGEDCCTSECTNICDCAKTKIKATFOCALGCCCKQPK 120
Qy 461 AGWVRPAKWEWLPPEWNGKSNWPMWQVNGWPHHGGKHLWMTGPTLQEQMTLW 520
Db 121 AGWVRPAKDECDLPEMCGKSGNCPDRFOVNGFPCHGGKHLWMTGPTLQEQMTLW 180
Qy 521 GPGTEVAKWSYNNRNEGSKYGYRRVWMTLIPKANTWMTGKLVWQGGSNLWPKGRIV 580
Db 181 GPGTEVADKSCYNNRNEGSKYGYRRVDDTLIPKANDTMCGLKFCQGGSNLWPKGRIV 240
Qy 581 TMLTWKTPPEWNTSQTGIGWANGTKGWNKVINAEWVIEKAYKSTNWSKKGHAYW 640
Db 241 TELCKTDPEDTSQTGIGWANGTKGWNKVINAEWVIEKAYKSTNWSKKGHAYW 300
Qy 641 HELQWEEGPIPPMWMSSVYVWMSIVVGLWPMVAVIWWVWVIRHQSREKQKQW 700
Db 301 HELQWEEGPIPPMWMSSVYVWMSIVVGLWPMVAVIWWVWVIRHQSREKQKQW 360
Qy 701 PLSTTGTRPHKQKRPQWKAQVQPMESOMKPHVYVWLVVEGNEPPASHKTNALPPTVW 760
Db 361 PLSTTGTRPHKQKRPQWKAQVQPMESOMKPHVYVWLVVEGNEPPASHKTNALPPTVW 420
Qy 761 KNPMTSTPKDSNPKA 775
Db 421 KNPMTSTPKDSNPKA 435

RESULT 9

AA64681
ID AAB64681 standard; protein; 435 AA.
AC AAB64681;
XX AAB64681;
DT 23-MAR-2001 (first entry)
XX Human secreted protein sequence encoded by gene 16 SEQ ID NO:75.
XX Human; secreted protein; diagnosis; cytostatic; antirheumatic;
KW antiarthritic; dermatological; cardiant; antiinflammatory; anti-ulcer;
KW gastrointestinal; solid tumour; rheumatoid arthritis; psoriasis;
KW diabetic retinopathy; myocardial angiogenesis; Crohn's disease; ulcer.
XX Homo sapiens.
OS Homo sapiens.
XX WO200077237-A1.
XX 21-DEC-2000.
XX 01-JUN-2000; 2000WO-US014928.
XX 11-JUN-1999; 99US-0138633P.
XX (HOMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX Rosen CA, Ruben SM, Komatsoulis GA;
XX WPI; 2001-071280/08.
XX N-PSDB; AAF33052.
XX Nucleic acids encoding 49 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating diseases such as tumors,
PT rheumatoid arthritis, psoriasis and diabetic retinopathy.
XX Claim 11; Page 465-467; 520pp; English.
XX The polynucleotide sequences given in AAF33037 to AAF33095 encode the
CC human secreted proteins given in AAB6466 to AAB64714. AAB64715 to
CC AAB64771 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are given in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissues
CC and cells the genes are expressed in. Examples of activities include:

CC cytostatic; antirheumatic; antiarthritic; dermatological; cardiant;
CC antiinflammatory; gastrointestinal; and anti-ulcer. The polynucleotides
CC and polypeptides can be used in the prevention, treatment and diagnosis
CC of diseases associated with inappropriate polypeptide expression.
CC Disorders that may be treated or prevented include solid tumours,
CC rheumatoid arthritis, psoriasis, diabetic retinopathy, myocardial
CC angiogenesis, Crohn's disease and ulcers. The polynucleotides and their
CC complementary sequences may also be used as DNA probes in diagnostic
CC assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate
CC the presence of similar nucleic acid sequences in samples, and therefore
CC which patients may be in need of restorative therapy. The polypeptides
CC may also be used as antigens in the production of antibodies against the
CC polypeptide and in assays to identify modulators (agonists and
CC antagonists) of polypeptide expression and activity. The anti-polypeptide
CC antibodies and antagonists may also be used to down regulate expression
CC and activity. AAF33028 to AAF33036 and AAB64665 represent sequences used
CC in the exemplification of the present invention
XX Sequence 435 AA;
SQ
Query Match 34.1%; Score 1605; DB 4; Length 435;
Best Local Similarity 79.1%; Pred. No. 4.9e-122;
Matches 344; Conservative 20; Mismatches 71; Indels 0; Gaps 0;
Qy 341 MCHNKGMMHMYWYKWPSTIWMKALSWYIPTMSSWSRLSYKWKWELGNLWNAFL 400
Db 1 MCHNKGMMHMYWYKWPSTIWMKALSWYIPTMSSWSRLSYKWKWELGNLWNAFL 60
Qy 401 PTWIIISTPIGQNLVEMGEDCCTSECTNICDCAKTKIKATFOCALGCCCKQPK 460
Db 61 PTWIIISTPIGQNLVEMGEDCCTSECTNICDCAKTKIKATFOCALGCCCKQPK 120
Qy 461 AGWVRPAKWEWLPPEWNGKSNWPMWQVNGWPHHGGKHLWMTGPTLQEQMTLW 520
Db 121 AGWVRPAKDECDLPEMCGKSGNCPDRFOVNGFPCHGGKHLWMTGPTLQEQMTLW 180
Qy 521 GPGTEVAKWSYNNRNEGSKYGYRRVWMTLIPKANTWMTGKLVWQGGSNLWPKGRIV 700
Db 301 GPGTEVADKSCYNNRNEGSKYGYRRVDDTLIPKANDTMCGLKFCQGGSNLWPKGRIV 360
Qy 701 PLSTTGTRPHKQKRPQWKAQVQPMESOMKPHVYVWLVVEGNEPPASHKTNALPPTVW 760
Db 361 PLSTTGTRPHKQKRPQWKAQVQPMESOMKPHVYVWLVVEGNEPPASHKTNALPPTVW 420
Qy 761 KNPMTSTPKDSNPKA 775
Db 421 KNPMTSTPKDSNPKA 435
RESULT 10
ABU60627
ID ABU60627 standard; protein; 435 AA.
XX ABU60627;
AC ABU60627;
XX ABU60627;
DT 02-MAY-2003 (first entry)
XX Human novel ADAM protein 3.
XX Human; enzyme; ADAM; a dystrophin and metalloprotease; cancer; stroke;
KW neural disease; cardiovascular disease; heart disease; atherosclerosis;
KW angina; reproductive disease; gastrointestinal disease; Crohn's disease;
KW endocrine disease; renal disease; central nervous system disease; AIDS;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; asthma;
KW schizophrenia; inflammatory disease; immunological disease; osteoporosis;

KW rheumatoid arthritis; inflammatory bowel disease; bone formation disease;
 KW musculo-skeletal disease; angiogenic disease; blood related disorder;
 KW respiratory neoplasia; thrombosis; gene therapy.

XX Homo sapiens.

XX US2002173640-A1.

XX 21-NOV-2002.

XX 19-APR-2002; 2002US-00125452.

XX 27-MAY-1999; 99US-0136388P.

XX 09-JUL-1999; 99US-0142230P.

XX 28-JAN-2000; 2000US-0178717P.

XX 25-MAY-2000; 2000WO-US014308.

XX 21-SEP-2000; 2000US-0234222P.

XX 16-NOV-2000; 2000US-00712307.

XX 19-SEP-2001; 2001US-00955504.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Ni J, Hastings GA, Shi Y, Wei P;

XX N-PSDB; ABX90528.

XX New isolated ADAM (a disintegrin and metalloprotease domain) polypeptide

XX and encoding polynucleotide useful for diagnosing and treating ADAM-

XX related pathologies e.g. cancer, blood disorders, inflammatory and immune

XX disorders and thrombosis.

XX Claim 1; Page 128-129; 147pp; English.

XX The invention relates to a new isolated polypeptide, comprising a

XX sequence at least 95% identical to any one of 7 fully defined sequences

XX given in the specification, or its fragment, epitope or variant. The

XX polypeptides are ADAM proteins (a dystrophin and metalloprotease). Also

XX included are the nucleic acids encoding the ADAM proteins, a vector

XX comprising the isolated nucleic acid, a host cell comprising the vector,

XX a recombinant host cell comprising the nucleic acid operably linked to a

XX heterologous regulating element which controls gene expression, an

XX isolated antibody that specifically binds to the ADAM protein,

XX identifying a binding partner to ADAM (comprising contacting ADAM with a

XX binding partner, and determining whether the binding partner affects an

XX activity of the polypeptide) and screening for molecules which modify

XX activities of ADAM (comprising contacting the polypeptide with a compound

XX suspected of having agonist or antagonist activity, and assaying for

XX activity of the polypeptide). The methods and compositions of the present

XX invention are useful for diagnosing, preventing, ameliorating and/or

XX treating diseases associated with the aberrant expression or activity of

XX the ADAM polypeptide, such as neural, cardiovascular (e.g. heart disease,

XX atherosclerosis, stroke, angina), reproductive, gastrointestinal (e.g.

XX Crohn's disease), endocrine, renal, central nervous system (e.g.

XX Alzheimer's, Parkinson's, Huntington's, schizophrenia), inflammatory

XX immunological (e.g. AIDS, rheumatoid arthritis, asthma, inflammatory

Qy 401 PTWIIISTPIGNOVLVGEHWWGCTSEBWTNIWWAKTWIKIKATQWALGEWWEKQK 460

Db 61 PTDIISTPICGNQVWGEDCDDCTSECTNICCDATCKIKATFQCALGSCCEKQPK 120

Qy 461 AGVVRDAKWEWMLPEMNGKSGNPNWVRWVONGWPNWHGSGHMLMGTWPTIQEOWTELW 520

Db 121 AGVCRPAKDECDLPENCKSGKNCDDPDRFQVNGFFCHHGKSGHCLMGTCTIQEOWTELW 180

Qy 521 GPGTEVANKWYNRNEGSKYGYVRRVWVTLIPKAWNTMWSKGLWQGGSNLWPKGRIV 580

Db 181 GPGTEVADKSCYNRNEGSKYGYCRVDDTLIPCKANDTMOGKLFQCGSDNLWPKGRIV 240

Qy 581 TWLTKTWMPWNTSOEIGMVGANGTKGNKWNINAEVWVWISKAYKSTNWSKWKSHAYVM 640

Db 241 TFLTKCTFDPEDTSOIEIGMVGANGTKGNKWNINAEVWVWISKAYKSTNWSKWKSHAYVCD 300

Qy 641 HELQWMEGWIPEMNMWSSVWHSIVVGVLMNAVIVVYVAVVIRHQSREKQKQOR 700

Db 301 HELQCCGEGWIPEDCDSSVVFHFSIVGVLFPMVIFVAVVIRHQSREKQKQOR 360

Qy 701 PLSTTGTRPHKQKRPQWKAQVQEMSKOMPHVWLVPEGNEPPASWHKTNALPPTVM 760

Db 361 PLSTTGTRPHKQKRPQWKAQVQEMSKOMPHVWLVPEGNEPPASWHKTNALPPTVF 420

Qy 761 KNPWSTPKNSPKA 775

Db 421 KNPWSTPKNSPKA 435

XX ID AAB64745 standard; protein; 367 AA.

XX AC AAB64745;

XX DT 23-MAR-2001 (first entry)

XX DE Human secreted protein sequence encoded by gene 16 SEQ ID NO:139.

XX KW Human; secreted protein; diagnosis; cytostatic; antirheumatic;

XX KW antiarthritic; dermatological; cardiac; antiinflammatory; anti-ulcer;

XX KW gastrointestinal; solid tumour; rheumatoid arthritis; psoriasis;

XX KW diabetic retinopathy; myocardial angiogenesis; Crohn's disease; ulcer.

XX OS Homo sapiens.

XX WO2000077237-A1.

XX 21-DEC-2000.

XX 01-JUN-2000; 2000WO-US014928.

XX 11-JUN-1999; 99US-0138633P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (ROSE/) ROSEN C A.

XX Rosen CA, Ruben SM, Komatsoulis GA;

XX WPI; 2001-071280/08.

XX Nucleic acids encoding 49 human secreted polypeptides, useful for

XX preventing, diagnosing and/or treating diseases such as tumors,

XX rheumatoid arthritis, psoriasis and diabetic retinopathy.

XX Disclosure; Page 505-506; 520pp; English.

XX The polynucleotide sequences given in AAF33037 to AAF33085 encode the

XX human secreted proteins given in AAB6466 to AAB64714. AAB64715 to

XX AAB64771 represent human secreted polypeptide sequences and proteins

XX homologous to them, which are given in the exemplification of the present

XX invention. Human secreted proteins have activities based on the tissues

KW

KW

KW

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Qy

Db

1

4

CC and cells the genes are expressed in. Examples of activities include:
 CC cystic; antirheumatic; antiarthritic; dermatological; cardiant;
 CC antiinflammatory; gastrointestinal; and anti-ulcer. The polynucleotides
 CC and polypeptides can be used in the prevention, treatment and diagnosis
 CC of diseases associated with inappropriate polypeptide expression.
 CC Disorders that may be treated or prevented include solid tumours,
 CC rheumatoid arthritis, psoriasis, diabetic retinopathy, myocardial
 CC angiogenesis, Crohn's disease and ulcers. The polynucleotides and their
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate
 CC the presence of similar nucleic acid sequences in samples, and therefore
 CC which patients may be in need of restorative therapy. The polypeptides
 CC may also be used as antigens in the production of antibodies against the
 CC polypeptide and in assays to identify modulators (agonists and
 CC antagonists) of polypeptide expression and activity. The anti-polypeptide
 CC antibodies and antagonists may also be used to down regulate expression
 CC and activity. AAF33028 to AAF33036 and AAB6465 represent sequences used
 CC in the exemplification of the present invention

XX SQ Sequence 367 AA;

Query Match 27.2%; Score 1284; DB 4; Length 367;
 Best Local Similarity 76.8%; Pred. No. 5.2e-96;
 Matches 282; Conservative 18; Mismatches 67; Indels 0; Gaps 0;
 QY 340 EMGNHNGHWHYSHKWBSTIWMKALSWIPTWSSSRSLSYKWEKLSNWLWNP 399
 Db 1 EMGNHNGHWHYSHKWBSTIWMKALSWIPTWSSSRSLSYKWEKLSNWLWNP 60
 QY 400 LPTWITIPICGNQIVEMGECCTCTSECTNICCDATCKIKATFQALGECCEKQFK 459
 Db 61 LPTWITIPICGNQIVEMGECCTCTSECTNICCDATCKIKATFQALGECCEKQFK 120
 QY 460 KAGMYRPAKWEWLPWMMNGSGNWPWRWQVNGWPHHGHGWLMTPTLQEQWTEL 519
 Db 121 KAGMYRPAKWEWLPWMMNGSGNWPWRWQVNGWPHHGHGWLMTPTLQEQWTEL 180
 QY 520 WPGTEVAKSWYRNEGSKYGYRRVWTLIPKANWTKWKLWQGGSNLWPKGRI 579
 Db 191 WPGTEVAKSWYRNEGSKYGYRRVWTLIPKANWTKWKLWQGGSNLWPKGRI 240
 QY 580 VTMLTWTWPEWTSQELGWANGTKGWNKWINAEVWIEKAYKSTNWSKKGHAWV 639
 Db 241 VTMLTWTWPEWTSQELGWANGTKGWNKWINAEVWIEKAYKSTNWSKKGHAWV 300
 QY 640 WHELOWQWEGWIPWMMWSSVVMHWSIVVGLVPMVIVVWVAMVIRHQSRKQKQW 699
 Db 301 WHELOWQWEGWIPWMMWSSVVMHWSIVVGLVPMVIVVWVAMVIRHQSRKQKQW 360
 QY 700 RPLSTTG 706
 Db 361 RPLSTTG 367

RESULT 12

AB53474 ID AAB53474 standard; protein; 233 AA.

XX AC AAB53474;

XX 09-MAR-2001 (first entry)

DE Human colon cancer antigen protein sequence SEQ ID NO:1014.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW identification; cytostatic; cardioactive; neuroprotective; vulnary;
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;
 KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
 KW neural disorder; immune system disorder; muscular disorder;
 KW reproductive disorder; gastrointestinal disorder; renal disorder;
 KW infectious disease; cardiovascular disorder.
 XX Homo sapiens.

XX PN WC200055351-A1.
 XX PD 21-SEP-2000.
 XX PF 08-MAR-2000; 2000WO-US005883.
 XX PR 12-MAR-1999; 99US-0124270P.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PI Rosen CA, Ruben SM;
 XX WPI: 2000-587534/55.
 XX N-PSDB; AAC98231.
 XX Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer.

XX Claim 11; Page 1599-1600; 2104pp; English.

XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
 CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
 CC vulnary, nephrotropic, antiinfective and antibacterial activities, and
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins may
 CC also be used to prevent diseases such as neural disorders, immune system
 CC disorders, muscular disorders, reproductive disorders, gastrointestinal
 CC disorders, wounds, renal disorders, infectious diseases, and
 CC cardiovascular disorders. AAC98764 to AAC98772 and AAB54007 represent
 CC sequences used in the exemplification of the present invention

XX Sequence 233 AA;

Query Match 18.1%; Score 854; DB 3; Length 233;
 Best Local Similarity 82.5%; Pred. No. 2.8e-61;
 Matches 179; Conservative 9; Mismatches 29; Indels 0; Gaps 0;

QY 559 TWMGKLMWQGGSNLWPKGRIVTMLTWTWPEWTSQELGWANGTKGWNKWINAEV 618
 Db 17 TWMGKLMWQGGSNLWPKGRIVTMLTWTWPEWTSQELGWANGTKGWNKWINAEV 76

QY 619 WIEKAYKSTNWSKKGHAWVWHELOWQWEGWIPWMMWSSVVMHWSIVVGLVPMV 678
 Db 77 WIEKAYKSTNWSKKGHAWVWHELOWQWEGWIPWMMWSSVVMHWSIVVGLVPMV 136

QY 679 VYVAMVIRHQSRKQKQWRLSTTGTRPHKQKRPQMKVAVOPQESQMKPHVYMLP 738
 Db 137 VYVAMVIRHQSRKQKQWRLSTTGTRPHKQKRPQMKVAVOPQESQMKPHVYMLP 196

QY 739 VEGNEPPASWKKTNALDPTWKKWNPMTPTKSNPKA 775

Db 197 VEGNEPPASWKKTNALDPTWKKWNPMTPTKSNPKA 233

RESULT 13

AAW90853

ID AAW90853 standard; protein; 201 AA.

XX AC AAW90853;

DT 28-JUN-2000 (first entry)

XX Human ADAM protein fragment #2.

DE ADAM protein; human; A disintegrin and metalloprotease; diagnosis;
 KW drug development; intervertebral hernia; sciatica; pulmonary fibrosis;
 KW

KW diabetic nephropathy; hepatic fibrosis; glomerulitis; osteopetrosis.
XX Homo sapiens.
XX WO200014227-A1.
XX 16-MAR-2000.
XX
PF 02-SEP-1999; 99WO-JP004766.
XX
FR 03-SEP-1998; 98JP-00250115.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
EI Yoshimura K, Hikichi Y, Nishimura A;
XX WP1; 2000-271056/23.
XX
PT Novel protein belong to A disintegrin and metalloprotease family, with
PT protease activity and extracellular matrix digesting enzyme activity, for
PT gene diagnosis and developing drugs for treating e.g. sciatica and
PT glomerulitis.
XX
PS Claim 23; Page 95-96; 109pp; Japanese.
XX
CC This invention describes a novel human protein (I) which belongs to the
CC ADAM (A disintegrin and metalloprotease) protein family. The protein,
CC peptide fragment and antibody are useful for gene diagnosis and in the
CC development of drugs to prevent or treat intervertebral hernia, sciatica,
CC glomerulitis, diabetic nephropathy, hepatic fibrosis, pulmonary fibrosis
CC or osteopetrosis. This sequence represents a human ADAM protein fragment
CC which is described in the method of the invention.
XX
SQ Sequence 201 AA;

Query Match 17.3%; Score 817; DB 3; Length 201;
Best Local Similarity 84.6%; Pred. No. 2.4e-58;
Matches 170; Conservative 12; Mismatches 19; Indels 0; Gaps 0;

QY 199 VQEHKEITYYLVLNKGWFKRYNENQWIRKRVEMANYVNMVYKLLNTHVALVGMETWT 258
Db 1 VQEHKEITYYLVLNKGWFKRYNENQWIRKRVEMANYVNMVYKLLNTHVALVGMETWT 60

QY 259 MKWIKITPNASWTLNENKWSGSLSRKRRHWIAQLITATLACTTGLAWMTWSPY 318
Db 6- DXDKIKITPNASFLENFSKWSGSLSRKRRHWIAQLITATLACTTGLAWMTWSPY 120

QY 319 SVGVVQHSNLLRVAGTMAHEMGNWGWFWWYSGKWPSTIIVNWKALSWYIPTWSSW 378
Db 121 SVGVVQHSNLLRVAGTMAHEMGNWGWFWWYSGKWPSTIIVNWKALSWYIPTWSSW 180

QY 379 SRLSYKWKWENKLSNWLWNP 399
Db 181 SRLSYKWKWENKLSNWLWNP 201

RESULT 14
AAB50938
ID AAB50938 standard; protein; 208 AA.
XX
XX AAB50938;
XX
XX 20-MAR-2001 (first entry);
XX
XX ADAM protein #4.
XX ADAM; disintegrin; metalloprotease; food additive; breast cancer;
KW ovarian.
KW Homo sapiens.
OS
FN WO200073223-A2.
XX

PD 07-DEC-2000.
XX
XX 25-MAY-2000; 2000WO-US014308.
XX
PR 27-MAY-1999; 99US-0136388P.
PR 09-JUL-1999; 99US-0142930P.
PR 28-JAN-2000; 2000US-0178717P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Wei P, Ni J, Hastings GA, Shi Y;
XX WPI; 2001-016507/02.
XX
XX Seven nucleic acid molecules encoding ADAM polypeptides containing a
PT disintegrin and metalloprotease domain, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases.
XX
PS Claim 11; Page 277-278; 287pp; English.
XX
CC The present invention relates to seven members of the ADAMs (proteins
CC which contain A Disintegrin And Metalloprotease domain) protein family.
CC The ADAMs proteins and DNA may be used to treat disease, as a food
CC additive or preservative, for chromosome identification, as probes for
CC diagnosing a disorder related to the female reproductive system, the gene
CC particularly breast and/or ovary cancer. They are also useful in the gene
CC therapy of breast and ovarian cancer
XX
SQ Sequence 208 AA;

Query Match 17.3%; Score 815; DB 4; Length 208;
Best Local Similarity 80.1%; Pred. No. 3.6e-58;
Matches 173; Conservative 9; Mismatches 26; Indels 8; Gaps 1;

QY 560 MWGKLWQGGSWNLFWKGRIVTWLTWKTWPEWTSQEIGMWANGTKWGNKYWNAEWYW 619
Db 1 MCGKLFQCGSSDNLFWKGRIVTWLTWKTWPEWTSQEIGMWANGTKWGNKYWNAEWYW 60

QY 620 IEKAYKSTNWSKWKGHAVWVHQLQWMEGWIIPWMMWSSVVMWWSIVVGLFPMVITW 679
Db 61 IEKAYKSTNWSKWKGHAVWVHQLQWMEGWIIPWMMWSSVVMWWSIVVGLFPMVITW 120

QY 680 VVWVAVIRHQSREKQKQVRLSTGTGPHKQKQKQVQKQVQKQVQKQVQKQVQKQV 739
Db 121 VVWVAVIRHQSREKQKQVRLSTGTGPHKQKQKQVQKQVQKQVQKQVQKQVQKQV 172

QY 740 EGNPEPPASWHKWTNALPPTVMKNPMTSEKWSNPXA 775
Db 173 EGNPEPPASWHKWTNALPPTVMKNPMTSEKWSNPXA 208

RESULT 15
ABU60628
ID ABU60628 standard; protein; 208 AA.
XX
XX AC ABU60628;
XX
XX 02-MAY-2003 (first entry)
XX
XX Human novel ADAM protein 4.
XX
XX Human; enzyme; ADAM; a dystrophin and metalloprotease; cancer; stroke;
KW neural disease; cardiovascular disease; heart disease; atherosclerosis;
KW angina; reproductive disease; gastrointestinal disease; Crohn's disease;
KW endocrine disease; renal disease; central nervous system disease; AIDS;
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; asthma;
KW schizophrenia; inflammatory disease; immunological disease; osteoporosis;
KW rheumatoid arthritis; inflammatory bowel disease; bone formation disease;
KW musculo-skeletal disease; angiogenic disease; blood related disorder;
KW respiratory neoplasia; thrombosis; gene therapy.
XX
XX Homo sapiens.
OS

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM protein - protein search, using sw model

Run on: June 9, 2004, 18:03:31 ; Search time 55 Seconds
(without alignments)
51.372 Million cell ur

Title: US-10-726-148A-15 COPY 428 437

Perfect score: 73

Sequence: 1 EWTN1WWWAK 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300

```
Database : A_Genexp_29Jan04.*
1: Genesexp1980s.*
2: Genesexp1990s.*
3: Genesexp2000s.*
4: Genesexp2001s.*
5: Genesexp2002s.*
6: Genesexp2003a.*
7: Genesexp2003bs.*
8: Genesexp2004s.*
```

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	46	63.0	40	3	AA820691	Aab20691 Polymeric
2	44	60.3	701	4	AA870446	Abb70446 Drosophil
3	43	58.9	37	3	AA814596	Abb14596 HIV-1 iso
4	43	58.9	37	4	AA863923	Aag63923 Amino aci
5	43	58.9	37	4	AAU70699	Aau70699 HIV viral
6	43	58.9	47	3	AA814594	Aab14594 HIV-1 iso
7	43	58.9	47	4	AA863921	Aag63921 Amino aci
8	43	58.9	47	4	AAU70697	Aau70697 HIV viral
9	43	58.9	66	4	AB811235	Abg11235 Novel hum
10	43	58.9	68	4	AAU65573	Aau65573 Propionib
11	43	58.9	68	4	AB814735	Abb14735 Human ner
12	43	58.9	68	6	AB862092	Abm62092 Propionib
13	43	58.9	138	4	AB805104	Abg05104 Novel hum
14	43	58.9	138	4	AB814739	Abb14739 Novel hum
15	43	58.9	138	4	AB814422	Abg14422 Novel hum
16	42	57.5	8	4	AA878562	AAH78562 HIV-1 gp
17	42	57.5	8	5	AD803156	Ade03156 Hybrid po
18	42	57.5	27	4	AA874349	Aag74349 Human col
19	42	57.5	96	2	AA679197	Aaw679197 Human sec
20	42	57.5	125	4	AAU30565	Aau30565 Novel hum
21	42	57.5	132	3	AA842895	Aab42895 Human ORF
22	42	57.5	143	2	AAV34588	Aay34588 Chlamydia
23	42	57.5	143	6	ABU26998	Abu26998 Protein e
24	42	57.5	282	6	ABU35120	Abu35120 Protein e
25	42	57.5	286	3	AA854774	Aag54774 Arabidops

99	39	53.4	100	3	AA041039	Ab041039 Human ORF	172	38.5	52.7	34	4	AA063567	Human bra
100	39	53.4	100	5	ABP09412	Abp09412 Human ORF	173	38.5	52.7	34	4	ABG58088	Human liv
101	39	53.4	108	7	ADC24675	Adc24675 HIV gp41	174	38.5	52.7	34	5	ABG45669	Human pep
102	39	53.4	110	4	AA065661	AA065661 Human rep	175	38.5	52.7	161	5	ABG60214	Human DIT
103	39	53.4	110	4	AB096565	Ab096565 Human tes	176	38.5	52.7	294	2	AAW69434	A. radiob
104	39	53.4	127	4	AAE01548	AAE01548 Human gen	177	38.5	52.7	294	2	AAW86275	Epichloro
105	39	53.4	127	4	AAE01579	AAE01579 Human gen	178	38.5	52.7	294	3	AA094505	A. radiob
106	39	53.4	127	5	ABG63830	Abg63830 Human alb	179	38.5	52.7	294	3	AA012222	Agrobacte
107	39	53.4	127	5	ABG63832	Abg63832 Human alb	180	38.5	52.7	794	4	AA090825	C glutami
108	39	53.4	127	6	ABR47985	Ab047985 Human sec	181	38	52.1	8	4	AA078560	HIV-1 gp
109	39	53.4	127	6	ABR47673	Ab047673 Human sec	182	38	52.1	8	4	AA078552	Hybrid po
110	39	53.4	127	6	ABR00034	Ab000034 Human gen	183	38	52.1	8	5	AA003146	Hybrid po
111	39	53.4	127	6	ABR00201	Ab000201 Human gen	184	38	52.1	9	3	AA090076	Enhancer
112	39	53.4	127	7	ABD91485	Abd91485 Human sec	185	38	52.1	9	3	AA090076	Enhancer
113	39	53.4	127	7	ABD91712	Abd91712 Human sec	186	38	52.1	9	4	AA006261	Anti-HIV
114	39	53.4	127	7	ADC74416	Adc74416 Human sec	187	38	52.1	9	4	AA006268	Anti-HIV
115	39	53.4	127	7	ADC74053	Adc74053 Human sec	188	38	52.1	9	4	AA076981	Retrovira
116	39	53.4	144	2	AAW05405	AAW05405 Mouse p53	189	38	52.1	9	4	AA076997	Retrovira
117	39	53.4	150	2	AAW69319	AAW69319 Anti-HIV	190	38	52.1	9	4	AB030351	Viral cor
118	39	53.4	157	5	ABU05465	Abu05465 M. tuberc	191	38	52.1	9	4	AB030355	Viral cor
119	39	53.4	160	6	ABP98171	Abp98171 Amino aci	192	38	52.1	9	5	AA003057	Hybrid po
120	39	53.4	161	5	ABU05901	Abu05901 M. tuberc	193	38	52.1	9	5	AA003073	Hybrid po
121	39	53.4	163	2	AAW69321	AAW69321 Anti-HIV	194	38	52.1	10	3	AA090085	Enhancer
122	39	53.4	171	3	AAU77261	AAU77261 Protein V	195	38	52.1	10	3	AA090087	Enhancer
123	39	53.4	173	3	AAU77259	AAU77259 Protein A	196	38	52.1	10	3	AA090096	Enhancer
124	39	53.4	194	7	ABD74557	Abd74557 Mycobacte	197	38	52.1	10	4	AAU06254	Anti-HIV
125	39	53.4	217	5	ABP51309	Abp51309 Human MDD	198	38	52.1	10	4	AAU06265	Anti-HIV
126	39	53.4	236	7	ADC25908	Adc25908 Human met	199	38	52.1	10	4	AAU06265	Anti-HIV
127	39	53.4	240	4	ABG07574	Abg07574 Novel hum	200	38	52.1	10	4	AA076990	Retrovira
128	39	53.4	243	2	AA022877	AA022877 SEQ ID NO	201	38	52.1	10	4	AA077001	Retrovira
129	39	53.4	243	3	AB015654	Ab015654 Protein e	202	38	52.1	10	4	AA076992	Retrovira
130	39	53.4	243	5	ABG68348	Abg68348 Envelope	203	38	52.1	10	4	AB030355	Viral cor
131	39	53.4	243	6	ABU57755	Abu57755 Human imm	204	38	52.1	10	4	AB030346	Viral cor
132	39	53.4	282	2	RAY32836	Ray32836 Woodchuck	205	38	52.1	10	5	AA003044	Viral cor
133	39	53.4	295	7	ADE08702	Ad08702 Novel pro	206	38	52.1	10	5	AA003066	Hybrid po
134	39	53.4	540	5	AAW49661	AAW49661 Rat SmGlu	207	38	52.1	10	5	AA003077	Hybrid po
135	39	53.4	715	2	AA055625	AA055625 HIV-1 gro	208	38	52.1	10	5	AA003077	Hybrid po
136	39	53.4	766	6	ADA53337	Ada53337 Human pro	209	38	52.1	20	5	ABJ15259	IGF relat
137	39	53.4	842	4	AA033965	AA033965 Human ABC	210	38	52.1	22	4	AA017488	Peptide #
138	39	53.4	842	5	ABP52102	Abp52102 Homo sapi	211	38	52.1	22	4	AB036509	Peptide #
139	39	53.4	842	5	AAE16764	AAE16764 Human tra	212	38	52.1	22	4	AA033008	Peptide #
140	39	53.4	842	7	AAE55514	AAE55514 Human pro	213	38	52.1	22	4	AB031304	Peptide #
141	39	53.4	843	4	AAU00011	AAU00011 Human ABC	214	38	52.1	22	4	AB021852	Protein #
142	39	53.4	843	4	AAU00010	AAU00010 Human ABC	215	38	52.1	22	4	AB021852	Protein #
143	39	53.4	877	2	AAW07205	AAW07205 HIV-1 gro	216	38	52.1	22	4	AA051356	Human liv
144	39	53.4	904	6	ABR52822	Ab052822 Protein s	217	38	52.1	22	4	AA051356	Human liv
145	39	53.4	912	2	AA082658	AA082658 Human mgl	218	38	52.1	22	4	AA051356	Human liv
146	39	53.4	912	2	AA072092	AA072092 Human mgl	219	38	52.1	22	5	AA039293	Human pep
147	39	53.4	912	5	AAE23757	AAE23757 Human met	220	38	52.1	24	3	AA051768	Human sec
148	39	53.4	912	6	ABP81846	Abp81846 Human met	221	38	52.1	30	5	AAU84535	HIV ENV s
149	39	53.4	912	7	AAE58170	AAE58170 Human pro	222	38	52.1	32	3	AA089807	Core poly
150	39	53.4	912	7	AAE58164	AAE58164 Rat Prote	223	38	52.1	32	4	AA078208	Core poly
151	39	53.4	912	7	AAE58168	AAE58168 Human pro	224	38	52.1	32	4	AB026666	Viral cor
152	39	53.4	912	7	AAE58166	AAE58166 Human pro	225	38	52.1	32	4	AB026666	Viral cor
153	39	53.4	1001	6	ABU89812	Abu89812 Novel hum	226	38	52.1	32	4	AAU013761	Viral DFI
154	39	53.4	1005	2	AAW93955	AAW93955 Human 53B	227	38	52.1	32	5	AA026686	Hybrid po
155	39	53.4	1005	6	ABU89810	Abu89810 Novel hum	228	38	52.1	35	3	AA089838	Core poly
156	39	53.4	1005	6	ABG71757	Abg71757 Human can	229	38	52.1	35	3	AA089838	Core poly
157	39	53.4	1005	6	ADA10593	Ada10593 Human can	230	38	52.1	35	5	AA026662	Hybrid po
158	39	53.4	1068	6	ABU89811	Abu89811 Novel hum	231	38	52.1	36	2	AA067699	DP-178 ho
159	39	53.4	1090	4	AA039292	AA039292 Human pol	232	38	52.1	36	2	AA067699	DP-178 ho
160	39	53.4	1096	4	AAW78805	AAW78805 Human pro	233	38	52.1	36	2	AAW17013	DP-178-li
161	39	53.4	1262	4	ABE70839	Ab070839 Drosophil	234	38	52.1	36	3	AA089635	Core poly
162	39	53.4	1467	5	AAU78283	AAU78283 Human apo	235	38	52.1	36	3	AA089635	Core poly
163	39	53.4	1609	5	AAU78282	AAU78282 Human apo	236	38	52.1	36	3	AA089675	Core poly
164	38.5	52.7	6	6	ABR46714	Ab046714 Staphyloc	237	38	52.1	36	3	AA089952	Core poly
165	38.5	52.7	20	2	AAW67506	AAW67506 IGF-1/IGF	238	38	52.1	36	3	AA089637	Core poly
166	38.5	52.7	20	4	AA063743	AA063743 Synthetic	239	38	52.1	36	3	AA089953	Core poly
167	38.5	52.7	20	5	ABE57660	Ab057660 IGFBP-3 b	240	38	52.1	36	3	AB014581	HIV-1 iso
168	38.5	52.7	20	7	ADU84826	Ad084826 Synthetic	241	38	52.1	36	4	AA063908	Amino aci
169	38.5	52.7	34	4	ABE42677	Ab042677 Peptide #	242	38	52.1	36	4	AA054787	HIV antiv
170	38.5	52.7	34	4	AA036490	AA036490 Peptide #	243	38	52.1	36	4	AA092246	Virus rel
171	38.5	52.7	34	4	AAW76382	AAW76382 Human bon	244	38	52.1	36	4	AA078038	Core poly

XX PN WO200159457-A2.
XX PD 16-AUG-2001.
XX PF C9-FEB-2001; 2001WO-US004030.
XX PR 10-FEB-2000; 2000US-0181543P.
XX PR 28-SEP-2000; 2000US-0235901P.
XX PA (PANA-) PANACOS PHARM INC.
XX PI Wild CT, Allaway GP;
XX DR WPI; 2001-522493/57.
XX PT Screening for inhibitors of viral entry structure formation by
XX PT selectively triggering the formation of one or more critical entry
XX PT intermediates in cell-surface-expressed viral envelope and probing for
XX PT its formation.
XX PS Disclosure; Page 34; 68pp; English.
XX CC The present sequence represents a fragment of a Human immunodeficiency
XX CC virus type 1 (HIV-1) gp41 protein. The peptide is used to raise
XX CC antibodies for use in the method of the invention. The specification
XX CC describes a method of screening for inhibitors of viral entry structure
XX CC formation. The method comprises contacting a viral envelope protein or
XX CC glycoprotein (e.g. gp41) with a triggering agent and a candidate compound
XX CC to form a mixture, and measuring the effect that the candidate compound
XX CC has on the formation of conformational intermediates. The effect of the
XX CC candidate compound can be measured by antibody binding to these
XX CC conformational intermediates. The compounds identified by the method are
XX CC useful as inhibitors for inhibiting or preventing viral infection and to
XX CC treat humans infected with HIV-1 or other viruses. This antiviral
XX CC compounds can also be used to inactivate viruses in body fluids, e.g.
XX CC blood or blood compounds used for therapeutic purposes. The assay is also
XX CC useful for detecting antibodies in virus-infected individuals or virus-
XX CC infected body fluids or tissues that inhibit entry-relevant
XX CC conformational changes in one or more viral envelope proteins or
XX CC glycoproteins. The presence of the antibodies in infected individuals or
XX CC samples is of prognostic value. (Updated on 11-SEP-2003 to standardise OS
XX CC field)
XX SQ Sequence 37 AA;
Query Match 58.9%; Score 43; DB 4; Length 37;
Best Local Similarity 62.5%; Pred. No. 31;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 EWTNINW 8
DB 28 KWTNINW 35
RESULT 5
AAU70699
ID AAU70699 standard; peptide; 37 AA.
XX AC AAU70699;
XX DT 14-FEB-2002 (first entry)
XX DE HIV viral envelope protein stabilising peptide #41.
XX KW Human; HIV-1; human immunodeficiency virus; gp41; glycoprotein 41; gp120;
XX KW anti-HIV; stabilising peptide; viral envelope protein; glycoprotein 120;
XX KW alpha-helical region; ectodomain.
XX OS Homo sapiens.
XX PN WO200170262-A2.
XX PD 13-JUL-2000.

PD 27-SEP-2001.
XX PF 15-MAR-2001; 2001WO-US008108.
XX PR 17-MAR-2000; 2000US-0189981P.
XX PA (PANA-) PANACOS PHARM INC.
XX PI Wild CT, Allaway GP;
XX DR WPI; 2001-626098/72.
XX PT Immunogenic composition for inhibiting HIV infection, comprises viral
XX PT envelope protein or its fragment exterior to viral membrane, a
XX PT stabilising peptide, and, optionally, viral cell surface receptor or its
XX PT fragment.
XX PS Disclosure; Page 39; 84pp; English.
XX CC The invention relates to methods of generating immunogens that elicit
XX CC neutralising antibodies which target regions of viral envelope proteins
XX CC such as the gp120/gp41 (glycoprotein 120/glycoprotein 41) complex of HIV-
XX CC 1 (human immunodeficiency virus 1). Sequences AAU70179-AAU70198 and
XX CC AAU70677-AAU70743 represent stabilising peptides modelling the alpha-
XX CC helical regions of the ectodomain of the HIV-1 transmembrane protein to
XX CC stabilise fusion-active intermediate structures, which can be used as
XX CC vaccine immunogens. Immunogenic compositions comprise a viral envelope
XX CC protein or its fragment exterior to the viral membrane, a stabilising
XX CC peptide to disrupt formation of structural intermediates necessary for
XX CC viral fusion and entry, and optionally, a viral cell surface receptor or
XX CC its fragment. The stabilising peptide is capable of associating with the
XX CC envelope protein or its fragment to form a stabilised, fusion active
XX CC structure. Antibody binding assays are used to determine the ability of
XX CC immunogen vaccines to generate an immune response to various forms of
XX CC envelope. Virus neutralisation assays can be used to characterise the
XX CC antibody response raised against HIV-1 gp41 domains. The sequences and
XX CC methods are useful for inhibiting HIV infection, for inducing an immune
XX CC response in an animal and for raising antibodies
XX SQ Sequence 37 AA;
Query Match 58.9%; Score 43; DB 4; Length 37;
Best Local Similarity 62.5%; Pred. No. 31;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 EWTNINW 8
DB 28 KWTNINW 35
RESULT 6
AAB14594
ID AAB14594 standard; peptide; 47 AA.
XX AC AAB14594;
XX DT 12-SEP-2003 (revised)
XX DT 24-NOV-2000 (first entry)
XX DE HIV-1 isolate SE.SE92809 gp41 C-helical domain peptide, SEQ ID NO:66.
XX KW HIV-1; gp41 C-helical domain;
XX KW gp41 transmembrane-proximal amphipathic alpha-helical segment;
XX KW core 6-helix bundle; viral entry inhibition; immunogenic; antibody;
XX KW humoral response; broad spectrum vaccine; anti-HIV;
XX KW envelope glycoprotein; prophylaxis; therapy.
XX OS Human immunodeficiency virus 1.
XX PN WO200040616-A1.
XX PD 13-JUL-2000.

```

PF 10-JAN-2000; 2000US-0000456.
XX
XX 08-JAN-1999; 99US-0115404P.
PR 07-JAN-2000; 2000US-00480336.
XX
XX (WILD/) WILD C T.
PA (WEIS/) WEISS C D.
XX
XX Wild CT, Weiss CD;
PI
XX WPI; 2000-465933/40.
DR
XX
XX Raising neutralizing antibody response to human immunodeficiency virus,
PT comprises administering a polypeptide capable of forming a stable coiled-
PT coil solution structure.
XX
XX Claim 13; Page 28; 97pp; English.
XX
XX Sequences AAB14532-B14534 and AAB14569-B14602 represent specifically
CC claimed peptides derived from the C-helical domain of the gp41 envelope
CC glycoprotein from a variety of HIV-1 isolates. The invention relates to
CC raising a neutralising antibody response to a broad spectrum of HIV
CC (human immunodeficiency virus) strains and isolates, comprising the
CC administration of a peptide which corresponds to or mimics highly
CC conserved portions of gp41 which are important in mediating the process
CC of viral entry into host cells. Such peptides can correspond to or mimic
CC the coiled coil solution structure of the N-helical domain (the heptad
CC repeat region), or can correspond or mimic the C-helical domain (the
CC transmembrane-proximal amphipathic alpha-helical segment), or the gp41
CC core 6-helix bundle, which is formed by the interaction of the N- and C-
CC helical domains of three gp41 proteins. The peptides can be administered
CC either singly or as a combination (particularly a combination of N-
CC helical and C-helical peptides), and can be multimerised. For example, N-
CC and C-helical domain peptides can be alternately linked together to form
CC a peptide which mimics the core 6-helix bundle. Administration of the
CC peptide(s) generates a humoral response, with the production of
CC antibodies against gp41 structures involved in viral entry. As these
CC portions of gp41 are well conserved, such antibodies may be effective
CC against a broad range of HIV strains and isolates. The peptide
CC compositions may be administered as a prophylactic or therapeutic vaccine
CC to generate antibodies which reduce or inhibit the ability of HIV to
CC infect uninfected cells. A composition comprising polyclonal or
CC monoclonal antibodies can be administered to reduce HIV infection of
CC uninfected cells. Antibodies raised against entry-relevant gp41
CC structures may also be used therapeutically and as tools to further
CC elucidate the mechanism of HIV cell entry. (Updated on 12-SEP-2003 to
CC standardise OS field)
XX
XX Sequence 47 AA;
SQ
Query Match 58.9%; Score 43; DB 3; Length 47;
Best Local Similarity 62.5%; Pred. No. 39;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 EWTNIWW 8
Db 38 KWTNLNWN 45
RESULT 7
AAG63921
ID AAG63921 standard; peptide; 47 AA.
XX
XX AAG63921;
AC
XX 11-SEP-2003 (revised)
DT 29-OCT-2001 (first entry)
XX
XX Amino acid sequence of a HIV-1 gp41 peptide fragment.
DE
XX HIV-1; gp41; viral entry; envelope protein; glycoprotein;
XX viral infection; antiviral.
XX

```

```

OS Human immunodeficiency virus 1.
XX
XX WO200159457-A2.
XX
XX 16-AUG-2001.
PD
XX
XX 09-FEB-2001; 2001WO-US004030.
PF
XX
XX 10-FEB-2000; 2000US-0181543P.
PR 28-SEP-2000; 2000US-0235901P.
XX
XX (PANA-) PANACOS PHARM INC.
PA
XX
XX Wild CT, Allaway GP;
PI
XX WPI; 2001-522493/57.
DR
XX
XX Screening for inhibitors of viral entry structure formation by
PT selectively triggering the formation of one or more critical entry
PT intermediates in cell-surface-expressed viral envelope and probing for
PT its formation.
XX
XX Disclosure; Page 34; 68pp; English.
XX
XX The present sequence represents a fragment of a Human immunodeficiency
CC virus type 1 (HIV-1) gp41 protein. The peptide is used to raise
CC antibodies for use in the method of the invention. The specification
CC describes a method of screening for inhibitors of viral entry structure
CC formation. The method comprises contacting a viral envelope protein or
CC glycoprotein (e.g. gp41) with a triggering agent and a candidate compound
CC to form a mixture, and measuring the effect that the candidate compound
CC has on the formation of conformational intermediates. The effect of the
CC candidate compound can be measured by antibody binding to these
CC conformational intermediates. The compounds identified by the method are
CC useful as inhibitors for inhibiting or preventing viral infection and to
CC treat humans infected with HIV-1 or other viruses. This antiviral
CC compounds can also be used to inactivate viruses in body fluids, e.g.
CC blood or blood compounds used for therapeutic purposes. The assay is also
CC useful for detecting antibodies in virus-infected individuals or virus-
CC infected body fluids or tissues that inhibit entry-relevant
CC conformational changes in one or more viral envelope proteins or
CC glycoproteins. The presence of the antibodies in infected individuals or
CC samples is of prognostic value. (Updated on 11-SEP-2003 to standardise OS
CC field)
XX
XX Sequence 47 AA;
SQ
Query Match 58.9%; Score 43; DB 4; Length 47;
Best Local Similarity 62.5%; Pred. No. 39;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 EWTNIWW 8
Db 38 KWTNLNWN 45
RESULT 8
AAU70697
ID AAU70697 standard; protein; 47 AA.
XX
XX AAU70697;
AC
XX 14-FEB-2002 (first entry)
DT
XX
XX HIV viral envelope protein stabilising peptide #39.
DE
XX Human; HIV-1; human immunodeficiency virus; gp41; glycoprotein 41; gp120;
XX anti-HIV; stabilising peptide; viral envelope protein; glycoprotein 120;
XX alpha-helical region; ectodomain.
XX
XX Homo sapiens.
OS
XX WO200170262-A2.
PN

```

XX PD 27-SEP-2001.
XX PF 15-MAR-2001; 2001WO-US008108.
XX PF 17-MAR-2000; 2000US-0189981P.
XX PF (PANA-) PANACOS PHARM INC.
XX PI Wild CT, Allaway GP;
XX PI WPI; 2001-626098/72.
XX PT Immunogenic composition for inhibiting HIV infection, comprises viral
XX PT envelope protein or its fragment exterior to viral membrane, a
XX PT stabilizing peptide, and, optionally, viral cell surface receptor or its
XX PT fragment.
XX PS Disclosure; Page 39; 84pp; English.
XX CC The invention relates to methods of generating immunogens that elicit
XX CC neutralising antibodies which target regions of viral envelope proteins
XX CC such as the gp120/gp41 (glycoprotein 120/glycoprotein 41) complex of HIV-
XX CC 1 (human immunodeficiency virus 1). Sequences AAU70179-AAU70198 and
XX CC AAU70677-AAU70743 represent stabilising peptides modelling the alpha-
XX CC helical regions of the ectodomain of the HIV-1 transmembrane protein to
XX CC stabilise fusion-active intermediate structures, which can be used as
XX CC vaccine immunogens. Immunogenic compositions comprise a viral envelope
XX CC protein or its fragment exterior to the viral membrane, a stabilising
XX CC peptide to disrupt formation of structural intermediates necessary for
XX CC viral fusion and entry, and optionally, a viral cell surface receptor or
XX CC its fragment. The stabilising peptide is capable of associating with the
XX CC envelope protein or its fragment to form a stabilised, fusion active
XX CC structure. Antibody binding assays are used to determine the ability of
XX CC immunogen vaccines to generate an immune response to various forms of
XX CC envelope. Virus neutralisation assays can be used to characterise the
XX CC antibody response raised against HIV-1 gp41 domains. The sequences and
XX CC methods are useful for inhibiting HIV infection, for inducing an immune
XX CC response in an animal and for raising antibodies
XX SQ Sequence 47 AA;
Query Match 58.9%; Score 43; DB 4; Length 47;
Best Local Similarity 62.5%; Pred. No. 39;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 EWTNIWW 8
DB 38 KWTNLWW 45
RESULT 9
ABG11235
ID ABG11235 standard; protein; 66 AA.
AC ABG11235;
XX 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #11226.
XX DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX DE Food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PF 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX PI WPI; 2001-639362/73.
XX DR N-PSDB; AAS75422.
XX XX New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID NO 41594; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX CC and in recombinant production of (II). The polynucleotides are also used
XX CC in diagnostics as expressed sequence tags for identifying expressed
XX CC genes. (I) is useful in gene therapy techniques to restore normal
XX CC activity of (II) or to treat disease states involving (II). (II) is
XX CC useful for generating antibodies against it, detecting or quantitating a
XX CC polypeptide in tissue, as molecular weight markers and as a food
XX CC supplement. (II) and its binding partners are useful in medical imaging
XX CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX CC involving aberrant protein expression or biological activity. The
XX CC polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
XX CC amino acid sequences of the invention. Note: The sequence data for this
XX CC patent did not appear in the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 66 AA;
Query Match 58.9%; Score 43; DB 4; Length 66;
Best Local Similarity 57.1%; Pred. No. 55;
Matches 4; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 WTNIIWW 8
DB 55 WAEAWWW 61
RESULT 10
AAU65573
ID AAU65573 standard; protein; 68 AA.
XX AC AAU65573;
XX AC 27-FEB-2002 (first entry)
XX DT Propionibacterium acnes immunogenic protein #26469.
XX DE SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
XX DE uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX DE inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX DE dermatological; osteopathic; neuroprotectant.
XX OS Propionibacterium acnes.
XX PN WO200181581-A2.
XX PD 01-NOV-2001.
XX PF 20-APR-2001; 2001WO-US012865.
XX PF 21-APR-2000; 2000US-0199047P.

```

PR 02-JUN-2000; 2000US-0208941P.
PR 07-JUL-2000; 2000US-0216747P.
XX (CORI-) CORIXA CORP.
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bvotia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
PI WPI; 2001-616774/71.
XX N-PSDB; AAS59677.
DR Propionibacterium acnes polypeptides and nucleic acids useful for
XX vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.
PT
XX Example 1; SEQ ID NO 26768; 1069pp; English.
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 68 AA;
XX
Query Match 58.9%; Score 43; DB 4; Length 68;
Best Local Similarity 57.1%; Pred. NO. 56;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 2 WTNTWVW 8
Db 33 WCVVWVW 39
RESULT 11
ABBI4735
ID ABBI4735 standard; protein; 68 AA.
XX
XX AC ABBI4735;
XX
XX 23-JAN-2002 (first entry)
XX
XX Human nervous system related polypeptide SEQ ID NO 3392.
XX
XX Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;
XX immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
XX antiparkinsonian; antisickling; antianaemic; antiarthritis; cancer;
XX antirheumatic; hepatotropic; cerebrotective; antiinflammatory;
XX antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
XX antiparasitic; cardiant; immune disorder; cardiovascular disorder;
XX neurological disease; infection; nephrotropic; gene therapy; vaccine.
XX
XX Homo sapiens.
XX
XX WO200159063-A2.
XX
XX 16-AUG-2001.
XX

```

```

PR 17-JAN-2001; 2001WO-US001334.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229503P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.

```

PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 20-OCT-2000; 2000US-0242221P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 01-DEC-2000; 2000US-0251160P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251983P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX MPI; 2001-541565/60.
 DR N-PSDB; ABA11061.
 XX

PT Nucleic acids encoding 3224 human nervous system antigen polypeptides, useful for preventing, diagnosing and/or treating nervous system cancers and metastases.
 PT
 XX Claim 11; SEQ ID NO 3392; 1701pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABA11004-ABA21534) and proteins (AB114678-AB118001) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 CC
 XX Sequence 68 AA;
 SQ

Query Match 58.9%; Score 43; DB 4; Length 68;
 Best Local Similarity 57.1%; Pred. No. 56;
 Matches 4; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 WTNIWW 8
 DB 48 WAGAWW 54
 |||||
 |||||

RESULT 12
 ABM62092
 ID ABM62092 standard; protein; 68 AA.
 XX AC ABM62092;
 XX 20-OCT-2003 (first entry)
 DT
 XX Propionibacterium acnes predicted ORF-encoded polypeptide #26768.
 DE
 XX Acne vulgaris; antiseborrheic; dermatological; antibacterial;
 KW immunostimulant; immune response; vaccine.
 XX OS Propionibacterium acnes.
 XX WO2003033515-A1.
 XX 24-APR-2003.
 PD
 XX 11-OCT-2002; 2002WO-US032727.
 PF
 XX 15-OCT-2001; 2001US-00978825.
 PR
 XX (CORI-) CORIXA CORP.
 PA
 XX Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
 PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
 PI Barth B, Vallieve-Douglass J;
 XX WPI; 2003-381789/36.
 DR N-PSDB; ACF64606.
 DR
 XX New Propionibacterium acnes polypeptides and polynucleotides encoding the polypeptide, useful for diagnosing, preventing or treating acne vulgaris, or for stimulating an immune response specific for a P. acnes protein.
 PT
 XX Example 1; SEQ ID NO 26768; 1481pp; English.
 PS
 XX

CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733) encoding a *Propionibacterium* acnes protein. The invention also relates to polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to immunogenic fragments of *P. acnes* polypeptides. The invention additionally encompasses expression vectors and host cells comprising a polynucleotide of the invention; antibodies against polypeptides of the invention; fusion proteins comprising a polypeptide of the invention; a method for stimulating an immune response specific for a *P. acnes* polypeptide and an isolated T cell population comprising T cells prepared via this method; a vaccine composition (comprising *P. acnes* polypeptides, polynucleotides, antibodies, fusion proteins, T cell populations, or antigen-presenting cells that express the polypeptide); a method and kit for detecting or determining the presence or absence of *P. acnes* in a patient; and a method for inhibiting the development of *P. acnes* in a patient. The *P. acnes* polypeptides, polynucleotides, antibodies, fusion proteins, T cell populations or antigen-presenting cells that express the polypeptides are useful for diagnosing, preventing or treating acne vulgaris, or for stimulating an immune response specific for a *P. acnes* protein. The polynucleotides can also be used as probes or primers for nucleic acid hybridisation. The vaccine composition is useful for the stimulation of an immune response against *P. acnes*, or for treating acne, and the kit is useful for performing a diagnostic assay. The present sequence represents a polypeptide predicted to be encoded by an ORF (open reading frame) contained within the *P. acnes* polynucleotides of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 68 AA;
Query Match 58.9%; Score 43; DB 6; Length 68;
Best Local Similarity 57.1%; Pred. No. 56;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 WTNITWVW 8
| : : : :
Db 33 WCVVWVWV 39

RESULT 13
ABG05104
ID ABG05104 standard; protein; 138 AA.

AC ABG05104;
XX
XX
DT 13-FEB-2002 (first entry)
DE
DE Novel human diagnostic protein #5095.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
CS Homo sapiens.

XX WO200175067-A2.
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX
XX 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX N-PSDB; AAS69291.
XX

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
XX

PT biodiversity.

XX Claim 20; SEQ ID NO 35463; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG0307 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 138 AA;

Query Match 58.9%; Score 43; DB 4; Length 138;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 NIWVW 8
| : : : :
Db 27 NIWVW 31

RESULT 14
ABG14739
ID ABG14739 standard; protein; 138 AA.

XX ABG14739;
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #14730.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
CS Homo sapiens.

XX WO200175067-A2.
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX
XX 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX N-PSDB; AAS78926.
XX

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX

PS Claim 20; SEQ ID NO 45098; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological actions in polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 138 AA;

Query Match 58.9%; Score 43; DB 4; Length 138;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NIWW 8
|
|
|
|
Db 27 NIWW 31

RESULT 15
ABG14422
ID ABG14422 standard; protein; 138 AA.

AC ABG14422;
DT 18-FEB-2002 (first entry)
DE Novel human diagnostic protein #14413.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
OS
XX WO200175067-A2.
PN
XX 11-OCT-2001.
PD
XX 30-MAR-2001; 2001WO-US008631.
PF
XX 31-MAR-2000; 2000US-00543217.
PR
XX 23-AUG-2000; 2000US-00649167.
FR
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI; 2001-639362/73.
DR
XX N-PSDB; AAS78609.
DR
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 44781; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological actions in polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 138 AA;

Query Match 58.9%; Score 43; DB 4; Length 138;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NIWW 8
|
|
|
|
Db 27 NIWW 31

RESULT 16
AAB78562
ID AAB78562 standard; peptide; 8 AA.

AC AAB78562;
DT 11-SEP-2003 (revised)
DT 19-APR-2001 (first entry)
DE HIV-1 gp 41 enhancer peptide #48.
XX Human immunodeficiency virus; HIV; gp 41; envelope protein; enhancer;
KW antiviral; anti-HIV; virucide; hepatotropic; antiinflammatory;
KW hybrid polypeptide; coiled-coil peptide interaction; viral infection.
XX fusion-related disorder; bacterial infection; viral infection.
XX Human immunodeficiency virus 1.
OS
XX WO200103723-A1.
PN
XX 18-JAN-2001.
PD
XX 10-JUL-2000; 2000WO-US018772.
PF
XX 09-JUL-1999; 99US-00350641.
FR
XX (TRIM-) TRIMERIS INC.
PA
XX Barney S, Guthrie KI, Merutka G, Anwer MK, Lambert DM;
PI
XX WPI; 2001-147136/15.
DR
XX New hybrid polypeptide, useful for preventing, treating and diagnosing
PT e.g. viral infections, comprises an enhancer peptide linked to a core
PT polypeptide.
XX
XX Disclosure; Fig 2B; 151pp; English.
PS
XX The present sequence is an enhancer peptide which may be linked to a core
CC polypeptide to form a novel hybrid polypeptide. The hybrid polypeptide

CC exhibits enhanced pharmacokinetic properties relative to those exhibited
CC by the core polypeptide when introduced into a living system. It is used
CC to increase the in vitro or ex vivo half-life of the core polypeptide.
CC The hybrid and core polypeptides can be used for modulating fusogenic
CC events and intracellular processes involving coiled-coil peptide
CC interactions. Other uses include preventing, treating and/or diagnosing
CC disorders involving fusion events (e.g. modulation of neurotransmitter
CC exchange and sperm-egg fusion), intracellular processes involving coiled-
CC coil peptides (e.g. bacterial infections) and viral infections that
CC involve cell-cell and/or virus-cell fusion (e.g. viral infections caused
CC by human immunodeficiency virus, respiratory syncytial virus, Epstein-
CC Barr virus, hepatitis B virus, Mason-Pfizer virus and polio virus). The
CC enhancer peptide sequence increases the half-life and reduces the
CC clearance rate of therapeutic peptides, which increases their efficacy
CC and minimises the incidence and severity of adverse side effects. In
CC addition, this increases the sensitivity of the diagnostic procedure in
CC which they are used. (Updated on 11-SEP-2003 to standardise OS field)

XX
SQ Sequence 8 AA;

Query Match 57.5%; Score 42; DB 4; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.4e+06;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 WTNLWW 8
|||:|
Db 1 WTNLWW 7

RESULT 17

ADE03156
ID ADE03156 standard; peptide; 8 AA.

XX
AC ADE03156;

XX
DT 29-JAN-2004 (first entry)

XX
DE Hybrid polypeptide pharmacokinetic enhancer peptide, SEQ ID No 1663.
XX
KW hybrid; enhancer; anti-fusogenic; antiviral; virucide; antidiabetic;
XX pharmacokinetic; fusogenic; insulin; diabetes.
XX
OS Unidentified.

XX
FH Key Location/Qualifiers

FT Modified-site 1 /note= "Residue is modified by acetyl group"

FT Modified-site 8 /note= "C-terminal amide"

XX
DN US6348568-B1.

XX
PD 19-FEB-2002.

XX
PF 20-MAY-1999; 99US-00315304.

XX
PR 20-MAY-1998; 98US-00082279.

XX
PA (TRIM-) TRIMERIS INC.

XX
PI Barney S, Guthrie KI, Merutka G, Anwer MK, Lambert DM;

XX
DR WPI; 2002-424396/45.

XX
PT New hybrid polypeptide for modulating fusogenic events for e.g. antiviral
PT activity, has enhancer peptide sequence derived from retroviral envelope
PT protein sequences linked to core polypeptide e.g. therapeutic protein.

XX
PS Disclosure; SEQ ID NO 1663; 70pp; English.

XX
CC The invention relates to a novel hybrid polypeptide comprising an
CC enhancer peptide sequence linked to a core polypeptide. The enhancer
CC peptide sequence comprises WQSEWQXI or WASLWENF. The invention also

CC includes novel peptides that exhibit anti-fusogenic activity, antiviral
CC activity and/or ability to modulate intracellular processes. The novel
CC hybrid polypeptide has virucide and antidiabetic activity. The enhancer
CC peptide sequence enhances pharmacokinetic properties of any core
CC polypeptide, for example, a polypeptide useful for the treatment or
CC prevention of a disease, or an imaging agent useful for imaging
CC structures in vivo. The core polypeptides and hybrid polypeptides are
CC useful for modulating fusogenic events and exhibit antifusogenic or
CC antiviral activity. The novel hybrid polypeptide is useful for decreasing
CC viral infection and modulating intracellular processes involving coiled-
CC coil peptide interactions. The novel hybrid polypeptide comprises insulin
CC or its fragment, so the core polypeptide is useful for ameliorating the
CC symptoms of forms of diabetes. The novel hybrid polypeptide is also
CC useful as a part of prognosis for preventing disorders including fusion
CC events and viral infection that involves cell-cell and/or virus-cell
CC fusion, and for diagnosis and in vivo imaging methods. This sequence
CC represents an enhancer peptide of the invention.

XX
SQ Sequence 8 AA;

Query Match 57.5%; Score 42; DB 5; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.4e+06;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 WTNLWW 8
|||:|
Db 1 WTNLWW 7

RESULT 18

AAG74349
ID AAG74349 standard; protein; 27 AA.

XX
AC AAG74349;

XX
DT 03-SEP-2001 (first entry)

XX
DE Human colon cancer antigen protein SEQ ID NO:5113.

XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma.

XX
OS Homo sapiens.

XX
PN WO200122920-A2.

XX
PD 05-APR-2001.

XX
PF 28-SEP-2000; 2000WO-US026524.

XX
PR 29-SEP-1999; 99US-0157137P.

XX
PR 03-NOV-1999; 99US-0163280P.

XX
PA (HUMA-) HUMAN GENOME SCI INC.

XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;

XX
DR WPI; 2001-235357/24.

XX
DR N-ESDB; AAH33780.

XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.

XX
PS Claim 11; Page 6826-6827; 9803pp; English.

XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome

CC that affect the activity of P by expressing inactive proteins or to
 CC supplement the patients own production of P. Additionally, N may be used
 CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
 CC into a host cell and culturing the cell to express the proteins. N and P
 CC can be used in the prevention, diagnosis and treatment of colorectal
 CC carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent
 CC sequences used in the exemplification of the present invention. N.B.
 CC pages 666 to 692 and page 7053 of the sequence listing were missing at
 CC time of publication, meaning no sequences are present for SEQ ID NO:1027
 CC to 1052, 7921 and 7922
 XX
 SQ Sequence 27 AA;

Query Match 57.5%; Score 42; DB 4; Length 27;
 Best Local Similarity 75.0%; Pred. No. 31;
 Matches 6; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 2 WTNIWWNA 9
 || |||||
 Db 12 WT--WWNA 17

RESULT 19
 AA067917
 ID AAW67917 standard; protein; 96 AA.

XX AAW67917;

AC AAW67917;

XX 25-MAR-1999 (first entry)

DE Human secreted protein encoded by gene 83 clone HHFGA11.

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX Homo sapiens.

XX Key Location/Qualifiers
 FH Misc-difference 47
 FT /label= unknown

FN WO9842738-A1.

XX 01-OCT-1998.

PD 19-MAR-1998; 98WO-US005311.

XX 21-MAR-1997; 97US-0041276P.

PR 21-MAR-1997; 97US-0041277P.

PR 21-MAR-1997; 97US-0041281P.

PR 21-MAR-1997; 97US-0042344P.

PR 30-MAY-1997; 97US-0048069P.

PR 30-MAY-1997; 97US-0048094P.

PR 30-MAY-1997; 97US-0048095P.

PR 30-MAY-1997; 97US-0048096P.

PR 30-MAY-1997; 97US-0048131P.

PR 30-MAY-1997; 97US-0048135P.

PR 30-MAY-1997; 97US-0048154P.

PR 30-MAY-1997; 97US-0048160P.

PR 30-MAY-1997; 97US-0048186P.

PR 30-MAY-1997; 97US-0048187P.

PR 30-MAY-1997; 97US-0048188P.

PR 30-MAY-1997; 97US-0048351P.

PR 30-MAY-1997; 97US-0048352P.

PR 30-MAY-1997; 97US-0048353P.

PR 30-MAY-1997; 97US-0048355P.

PR 30-MAY-1997; 97US-0050937P.

PR 05-AUG-1997; 97US-0054804P.
 PR 19-AUG-1997; 97US-0056370P.
 PR 02-OCT-1997; 97US-0060862P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX Young P, Greene JM, Ferrie AM, Ruben SM, Rosen CA, Duan R, Hu J;
 PI Florence KA, Olsen HS, Ebner R, Brewer LA, Moore PA, Shi Y;
 PI Lafleur DW, Ni J;
 XX WPI; 1999-070066/06.
 DR N-PSDB; AAX00722.
 XX
 PT New isolated human genes and the secreted polypeptides they encode -
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders.
 XX
 PS Claim 11; Page 336; 385pp; English.
 XX
 CC This sequence represents a secreted human protein encoded by the gene
 CC clone detailed in the descriptor line. The gene can be used to generate
 CC fusion proteins by linking to the gene to a human immunoglobulin Fc
 CC portion (e.g. AAX00602) for increasing the stability of the fused protein
 CC as compared to the human protein only. The invention relates to 87 novel
 CC genes and their fragments (nucleic acid sequences: AAX00611-X00724; amino
 CC acid sequences AAW67807-W68004) which are useful for preventing, treating
 CC or ameliorating medical conditions e.g. by protein or gene therapy. Also,
 CC pathological conditions can be diagnosed by determining the amount of the
 CC new polypeptides in a sample or by determining the presence of mutations
 CC in the new polynucleotides. Specific uses are described for each of the
 CC 87 polynucleotides, based on which tissues they are most highly expressed
 CC in (see AAX00611 for described uses)
 XX
 SQ Sequence 96 AA;

Query Match 57.5%; Score 42; DB 2; Length 96;
 Best Local Similarity 44.4%; Pred. No. 1.1e+02;
 Matches 4; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 WTNIWWNAK 10
 |||||
 Db 68 WCHPWMDR 76

RESULT 20
 AAU30565
 ID AAU30565 standard; protein; 125 AA.

XX AAU30565;

XX 18-DEC-2001 (first entry)

XX Novel human secreted protein #1056.

XX Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX Homo sapiens.

XX WO200179449-A2.

XX 25-OCT-2001.

XX 16-APR-2001; 2001WO-US008656.

XX 18-APR-2000; 2000US-00552929.

PR 26-JAN-2001; 2001US-00770160.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

```
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy.
XX
PS Claim 20; Page 311; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU3304 represent the amino acid sequences of novel human
CC secreted proteins of the invention
XX
SQ Sequence 125 AA;
Query Match 57.5%; Score 42; DB 4; Length 125;
Best Local Similarity 71.4%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 EWTNIW 7
DB 89 EWTNIW 95
RESULT 21
ID AAB42895
XX AAB42895 standard; protein; 132 AA.
AC AAB42895;
DT 08-FEB-2001 (first entry)
DE Human ORFX ORF2659 polypeptide sequence SEQ ID NO:5318.
XX
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antiinflammatory;
XX antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
XX antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX bone damage; cartilage damage; antiinflammatory disease; coagulation;
XX thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US008621.
XX
XX 31-MAR-1999; 99US-0127607P.
PR 02-APR-1999; 99US-0127636P.
PR 05-APR-1999; 99US-0127728P.
PR 30-MAR-2000; 2000US-00540763.
XX
```

```
(CURA-) CURAGEN CORP.
XX
XX Shimkets RA, Leach M;
XX
WPI; 2000-602362/57.
XX
XX N-PSDB; AAC77104.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease.
XX
XX Claim 11; Page 4489-4490; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnary;
XX antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
XX anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
XX cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
XX dermatological; immunosuppressive; antiinflammatory; antibacterial;
XX antiviral; antifungal; antirheumatic; antihypertensive; antianaemic. The
XX sequences can be used for determining the presence of or predisposition
XX to, or preventing or treating pathological conditions associated with an
XX ORFX-associated disorder. The nucleic acids can be used to express ORFX
XX proteins in gene therapy vectors. The proteins and nucleic acids may be
XX used to treat cancers, proliferative disorders, neurodegenerative
XX disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
XX diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
XX storage, systemic lupus erythematosus, severe combined immunodeficiency
XX (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
XX disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
XX cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
XX enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
SQ Sequence 132 AA;
Query Match 57.5%; Score 42; DB 3; Length 132;
Best Local Similarity 71.4%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 3 TNIWNA 9
DB 119 TSAWNA 125
RESULT 22
AAY34588
ID AAY34588 standard; protein; 143 AA.
XX
XX AAY34588;
AC
XX 17-OCT-2003 (revised)
XX 13-SEP-1999 (first entry)
XX
XX Chlamydia pneumoniae transmembrane protein sequence.
XX
XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
XX sinusitis; purulent otitis media; erythema nodosum; pharyngitis; vaccine;
XX neutralising epitope.
XX
XX Chlamydophila pneumoniae.
XX
XX WO9927105-A2.
XX
XX 03-JUN-1999.
XX
XX 20-NOV-1998; 98WO-IB001890.
XX
XX 21-NOV-1997; 97FR-00014673.
XX 04-NOV-1998; 98US-0107078P.
XX
XX (GSET ) GENSET.
```

XX Griffais R;
 XX WPI; 1999-357842/30.
 XX Genome sequence of Chlamydia pneumoniae.
 XX Page 613; Disclosure; 1912pp; English.
 XX AAY34584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae. C.
 CC pneumoniae causes respiratory disease such as pneumonia and bronchitis
 CC and is thought to be a contributing factor in heart disease, sarcoidosis,
 CC sinusitis, purulent otitis media, erythema nodosum or pharyngitis. The
 CC polypeptides encoded by the open reading frames of the C. pneumoniae
 CC genome (see AAY34584-Y35879) can be used in immunogenic compositions as
 CC vaccines. Vectors containing C. pneumoniae nucleotide sequences can also
 CC be used as immunogenic compositions, especially where the vector directs
 CC the expression of a neutralising epitope of C. pneumoniae. (Updated on 17
 CC -OCT-2003 to standardise OS field)
 XX SQ Sequence 143 AA;
 Query Match 57.5%; Score 42; DB 2; Length 143;
 Best Local Similarity 66.7%; Pred. No. 1.6e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 2 WTNIWW 7
 Db 117 WSNLWW 122
 RESULT 23
 ABU26978
 ID ABU26978 standard; protein; 143 AA.
 AC ABU26978;
 XX
 XX 23-OCT-2003 (revised)
 DT 19-JUN-2003 (first entry)
 XX Protein encoded by Prokaryotic essential gene #2505.
 DE Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX Chlamydia pneumoniae.
 XX WO200277183-A2.
 XX 03-OCT-2002.
 XX 21-MAR-2002; 2002WO-US009107.
 XX 21-MAR-2001; 2001US-00815242.
 XX 06-SEP-2001; 2001US-00948993.
 XX 25-OCT-2001; 2001US-0342923P.
 XX 08-FEB-2002; 2002US-00072851.
 XX 06-MAR-2002; 2002US-0362699P.
 XX (ELIT-) ELITRA PHARM INC.
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI; 2003-029926/02.
 DR N-PSDB; ACA30848.
 XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX Claim 25; SEQ ID NO 54902; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway;
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than S. aureus, S. typhimurium,
 CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
 CC standardise OS field)
 XX SQ Sequence 143 AA;
 Query Match 57.5%; Score 42; DB 6; Length 143;
 Best Local Similarity 66.7%; Pred. No. 1.6e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 2 WTNIWW 7
 Db 117 WSNLWW 122
 RESULT 24
 ABU35120
 ID ABU35120 standard; protein; 282 AA.
 AC ABU35120;
 XX
 XX 19-JUN-2003 (first entry)
 DT Protein encoded by Prokaryotic essential gene #20647.
 XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX Moraxella catarrhalis.
 XX WO200277183-A2.
 XX 03-OCT-2002.
 XX 21-MAR-2002; 2002WO-US009107.
 XX 21-MAR-2001; 2001US-00815242.
 XX 06-SEP-2001; 2001US-00948993.
 XX 25-OCT-2001; 2001US-0342923P.
 XX 08-FEB-2002; 2002US-00072851.
 XX 06-MAR-2002; 2002US-0362699P.
 XX (ELIT-) ELITRA PHARM INC.
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xc HH;
XX WPI; 2003-029926/02.
DR N-PSDB; ACA38990.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 63044; 1766pp; English.
PS
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 282 AA;
SQ
Query Match 57.5%; Score 42; DB 6; Length 282;
Best Local Similarity 62.5%; Pred. No. 3.1e-02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 2 WTNWVWNA 9
Db 72 WTWVFWNA 79
RESULT 25
RAG54774
ID AAG54774 standard; protein; 286 AA.
XX
XX AAG54774;
AC
XX
XX 18-OCT-2000 (first entry)
DT
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 69949.
DB
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX
XX EP1033405-A2.
FN
XX
XX 06-SEP-2000.
PD

XX
PF 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139889P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.

```
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143544P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147401P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 08-JUL-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 22-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 57.5%; Score 42; DB 3; Length 286;
Best Local Similarity 71.4%; Pred. No. 3.2e+02;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNIWM 7
Db 183 EWTETWM 189

RESULT 26
AAC42073
ID AAC42073 standard; protein; 383 AA.
XX AAC42073;
AC AAC42073;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 52424.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
```

PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 06-APR-1999; 99US-0127462P.
PR 08-APR-1999; 99US-0128234P.
PR 16-APR-1999; 99US-0128714P.
PR 19-APR-1999; 99US-0129845P.
PR 21-APR-1999; 99US-0130077P.
PR 23-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139450P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139763P.
PR 22-JUN-1999; 99US-0139817P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 26-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149910P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151388P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.

```
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159594P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 23-OCT-1999; 99US-0161404P.
PR 23-OCT-1999; 99US-0161405P.
PR 23-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 57.5%; Score 42; DB 3; Length 383;
Best local Similarity 71.4%; Pred. No. 4.3e+02;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNIW 7
Db 131 EWTNIW 137

RESULT 27
AAG42072
ID AAG42072 standard; protein; 432 AA.
AC AAG42072;
XX
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 52423.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX Hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EF1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99JS-0123548P.
XX 23-MAR-1999; 99JS-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139432P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140351P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
```

```
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144332P.
PR 20-JUL-1999; 99US-0144322P.
PR 21-JUL-1999; 99US-0144844P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 22-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR C2-AUG-1999; 99US-0146386P.
PR C2-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149358P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 26-AUG-1999; 99US-0151085P.
PR 27-AUG-1999; 99US-0151086P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155113P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.

PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 57.5%; Score 42; DB 3; Length 432;
Best Local Similarity 71.4%; Pred. No. 4.8e+02;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EWTNIWW 7
Db 180 EWTETWW 186

RESULT 28
AAG42071
ID AAG42071 standard; protein; 435 AA.
XX
AC AAG42071;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 52422.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
```


PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 22-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 25-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161932P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 57.5%; Score 42; DB 3; Length 435;
Best Local Similarity 71.4%; Pred. No. 4.8e+02;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EWTNIWN 7
Db 183 EWTETWN 189

RESULT 29
ABE60579
ID ABE60579 standard; protein; 1071 AA.
AC ABE60579;
XX
XX
DT 26-MAR-2002 (first entry)
DE Drosophila melanogaster polypeptide SEQ ID NO 8529.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
DR N-PSDB; ABL04682.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
XX Disclosure; SEQ ID NO 8529; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL31840-ABL26175) and the encoded proteins (ABBS7737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 1071 AA;

Query Match 57.5%; Score 42; DB 4; Length 1071;
Best Local Similarity 71.4%; Pred. No. 1.2e+03;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 4 NIWNAK 10
Db 1834 NEWNAK 1040

RESULT 30
AAP81185
ID AAP81185 standard; protein; 1594 AA.
XX
XX AAP81185;
XX
DT 25-MAR-2003 (revised)
DT 29-OCT-1992 (first entry)
XX
DE Sequence encoded by the lower reading frame of the peplomeric gene of
DE FIVP strain 79-1146.
XX
KW Vaccine; peplomeric protein; E2 gene.
XX
OS Feline infectious peritonitis virus.
XX
PN EP264979-A.
XX
PD 27-APR-1988.
XX
XX C1-SEP-1987; 87EP-00201657.
XX
PR C5-SEP-1986; 86NL-00002244.
XX
XX (DUIN) DUPHAR INT RES BV.
XX
PI Degroot RJ, Spaan WJM, Vanderzeij BAM;
XX WPI; 1988-114147/17.
DR N-PSDB; AAN81533.
XX
XX Gene for feline infectious peritonitis virus - and gene prod. useful as
XX antigenic protein for vaccine.
XX
XX Disclosure; Fig 1; 13pp; English.
XX
CC cDNA was prep'd. from FIVP strain 79-1146. AAN81533 gives the sequence of
CC the peplomeric gene in three reading frames. The top reading frame is an
CC open reading frame of 4356 nucleotides and has a coding capacity for a
CC precursor polypeptide having a mol. wt. of 160,470 (1452 AAs). The
CC beginning and the end of the E2 gene are indicated in the FT of AAN81533.
CC The first 18 N-terminal AAs have a strong hydrophobic character and
CC presumably comprise a cleavable signal peptide. The extreme carboxy-
CC terminal part comprises a region of 20 hydrophobic AAs, which presumably
CC serves as a transmembrane anchor. The FIVP peplomeric protein has 35
CC potential glycosylation sites, of which 22 are in the N-terminal part
CC (pos. 1-790) which corresponds to the S-part of the IBV E2 (see
CC AAP81183). N.B. IBV = infectious bronchitis virus. "X" in the AA sequence
CC denotes the translation of a stop codon. (Updated on 25-MAR-2003 to
CC correct PI field.)
XX
XX SQ Sequence 1594 AA;

Query Match 56.8%; Score 41.5; DB 1; Length 1594;
Best Local Similarity 50.0%; Pred. No. 2e+03;
Matches 6; Conservative 1; Mismatches 2; Indels 3; Gaps 1;

Qy 2 WTNJ---WWNAK 10
Db 1142 WYNIRCTWWNAK 1153

RESULT 31
AAE38380
ID AAE38380 standard; peptide; 20 AA.
XX AC AAE38380;
XX DT 20-NOV-2003 (first entry)
XX Erbb2 binding peptide (ESP)-4.
XX Erbb2 binding peptide; HER2; neu; breast cancer; protein therapy; EBP;
XX Erbb2.
XX Unidentified.
XX WO2003061559-A2.
XX 31-JUL-2003.
XX 15-OCT-2002; 2002WO-US032947.
XX 12-OCT-2001; 2001US-0329183P.
XX (JWVE-) UNIV VERMONT & STATE AGRIC COLLEGE.
XX Krag DN, Pero SC, Oligino L;
XX WPI; 2003-671426/63.
XX N-PSDB; AAD58063.
XX A composition for diagnosing, preventing or treating disorders
PT characterized by Erbb2 overexpression (e.g. breast cancer) comprises an
PT Erbb2 binding peptide that binds specifically to the extracellular domain
PT of Erbb2.
XX Claim 2; Page 12; 106pp; English.
XX The present invention relates to peptides and peptidomimetics that bind
CC to the extracellular domain of Erbb2 (also known as HER2 or neu).
CC Sequences of the invention are useful in the diagnosis, prevention and
CC treatment of disorders characterised by Erbb2 overexpression (e.g. breast
CC cancer). The invention is also useful in protein therapy. The present
CC sequence is Erbb2 binding peptide (EBP)
XX Sequence 20 AA;
SQ Query Match 56.2%; Score 41; DB 6; Length 20;
Best Local Similarity 62.5%; Pred. No. 31;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 EWTNIWW 8
Db 8 EWQNDWF 15
RESULT 32
AAM91572
ID AAM91572 standard; protein; 65 AA.
XX AC AAM91572;
XX DT 07-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen SEQ ID NO:19165.
DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis.
XX Homo sapiens.
OS
XX WO200157182-A2.
XX

PD 09-AUG-2001.
XX 17-JAN-2001; 2001WO-US001354.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 09-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.

PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0245474P.
 PR 08-NOV-2000; 2000US-0245475P.
 PR 08-NOV-2000; 2000US-0245476P.
 PR 08-NOV-2000; 2000US-0245477P.
 PR 08-NOV-2000; 2000US-0245478P.
 PR 08-NOV-2000; 2000US-0245523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249246P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2000US-0259678P.
 (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Barash SC, Ruben SM;
 PI WPI; 2001-483426/52.
 XX

DR N-PSDB; AAK64353.
 XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and metastasis.
 XX Claim 11; SEQ ID NO 19165; 3071pp + Sequence Listing; English.
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention
 XX
 SQ Sequence 65 AA;
 Query Match 56.2%; Score 41; DB 4; Length 65;
 Best Local Similarity 66.7%; Pred. No. 16+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 4 NIWVWA 9
 Db 13 SVWVWA 18
 RESULT 33
 ABR41226
 ID ABR41226 standard; protein; 77 AA.
 XX
 AC ABR41226;
 XX
 DT 02-JUN-2003 (first entry)
 XX
 DE Human DITHP receptor.
 XX
 KW Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
 KW cancer; cell proliferative disorder; autoimmune disorder;
 KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
 KW neurological disorder; gastrointestinal disorder; transport disorder;
 KW connective tissue disorder; drug screening; proteome analysis;
 KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
 KW disease model; toxicological testing; transcript imaging; receptor.
 XX
 OS Homo sapiens.
 XX
 PN WO200297031-A2.
 XX
 PD 05-DEC-2002.
 XX
 XX 27-MAR-2002; 2002WO-US010056.
 XX
 XX 28-MAR-2001; 2001US-0279619P.
 XX 29-MAR-2001; 2001US-0280067P.
 XX 29-MAR-2001; 2001US-0280068P.
 XX 16-MAY-2001; 2001US-0291280P.
 XX 17-MAY-2001; 2001US-0291829P.
 XX 17-MAY-2001; 2001US-0291849P.
 XX 19-JUN-2001; 2001US-0299428P.
 XX 20-JUN-2001; 2001US-0299776P.
 XX 20-JUN-2001; 2001US-0300001P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.

```
XX Daffo A, Jones AL, Tran AB, Bahl CR, Gietzen D, Chinn J;
PI Dufour GE, Hillman JL, Yu JY, Tudson O, Yap PE, Ameshey SR;
PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstein EH;
PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
PI Flores V, Marwaha R, Lo A, Lan RV, Urastka ME;
XX WPI: 2003-129518/22.
DR N-PSDB; ACC46169.
XX
PT Novel human diagnostic and therapeutic polypeptide useful for identifying
PT test compound which specifically binds to a polypeptide encoded by human
PT diagnostic and therapeutic polynucleotide, and to induce antibodies.
XX
XX Claim 27; SEQ ID NO 761; 591pp; English.
XX
XX The invention relates to novel human diagnostic and therapeutic
XX polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
XX proteins (DITHP; ABR41136-ABR41812). The invention also relates to
XX polynucleotide sequences at least 90% identical to the dithp cDNA
XX sequences of the invention; recombinant vectors, host cells and
XX transgenic organisms comprising a dithp nucleic acid sequence; the
XX recombinant production of DITHP proteins; antibodies specific for DITHP
XX proteins; microarrays comprising dithp nucleic acid sequences; methods of
XX detecting dithp nucleotide and protein sequences; methods of screening
XX for compounds which specifically bind a DITHP protein; and methods of
XX assessing the toxicity of test compounds using a dithp hybridisation
XX probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
XX diagnosis of a wide variety of conditions including cancer and other cell
XX proliferative disorders; autoimmune or inflammatory disorders; bacterial,
XX viral, fungal or parasitic infections; hormonal disorders; metabolic
XX disorders; neurological disorders; gastrointestinal disorders; transport
XX disorders; and connective tissue disorders. They may also be used to
XX screen for modulators of protein activity or gene expression. DITHP
XX proteins can additionally be used in analysis of the proteome of a tissue
XX or cell type and to induce antibodies. The dithp nucleic acids are
XX additionally useful in somatic or germline gene therapy of the disorders
XX mentioned above, as a source of antisense sequences, as a source of
XX probes and primers, in genotyping and identification of individuals, in
XX the generation of transgenic animal models of human disease or knock in
XX humanised animals, in toxicological testing, and in transcript imaging.
XX The present sequence represents a DITHP protein which has receptor
XX activity. Note: The sequence data for this patent did not form part of
XX the printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 77 AA;
Query Match 56.2%; Score 41; DB 6; Length 77;
Best Local Similarity 57.1%; Pred. No. 1.2e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 3 TNIWWWA 9
: ||||
Db 14 SQVWWWA 20
RESULT 34
ABG26422
ID ABG26422 standard; protein; 89 AA.
XX
AC ABG26422;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #26413.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN W0200175067-A2.
```

```
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX
XX 23-AUG-2000; 2000US-00649167.
XX
XX (HYSB-) HYSBQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX
XX N-PSDB; AAS90609.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 20; SEQ ID NO 56781; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG0010-ABG0377 represent novel human diagnostic
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 89 AA;
Query Match 56.2%; Score 41; DB 4; Length 89;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 IWWWA 9
: ||||
Db 2 IWWWA 6
RESULT 35
ABG83395
ID ABG83395 standard; protein; 104 AA.
XX
AC ABG83395;
XX
DT 06-SEP-2002 (first entry)
XX
DE Human evolution conservative protein CG11.44.
XX
KW Human; evolution conservative protein CG11.44; tumour; haemopathy;
KW HIV infection; development disturbance; immunological disease;
KW inflammation; cytostatic; anti-HIV.
XX
OS Homo sapiens.
XX
PN CN1333271-A.
```

2D 30-JAN-2002.
XX 07-JUL-2000; 2000CN-00117084.
XX 07-JUL-2000; 2000CN-00117084.
XX (SHAN-) SHANGHAI BIODOOR GENE DEV CO LTD.
XX Mao Y, Xie Y;
PI WPI; 2002-305589/35.
DR N-PSDB; ABN85127.
DR Novel polypeptide-human evolution conservative protein CGI 11.44 and
PT polynucleotide for encoding said polypeptide.
XX Claim 1; Page 27-28 (Disclosure); 34pp; Chinese.
PS The present sequence is the protein sequence for human evolution
CC conservative protein CGI11.44. The protein and its coding sequence are
CC useful for treating several diseases such as malignant tumours,
CC haemopathy, HIV infection, development disturbance, immunological disease
CC and various inflammations
XX Sequence 104 AA;
SQ

Query Match 56.2%; Score 41; DB 5; Length 104;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 WTNW 7
DB 93 WDNVW 98
||: |||
||: |||

RESULT 36
ABU60956
ID ABU60956 standard; protein; 109 AA.
XX
AC ABU60956;
XX
DT 08-MAY-2003 (first entry)
XX
DE Lung specific protein (LSP) #59.
XX
XX Human; gene therapy; vaccine; lung specific antigen; cancer diagnosis;
XX cancer monitoring; cancer staging; cancer imaging; lung cancer;
XX non-cancerous diseases of the lung; transgenic animal.
XX Homo sapiens.
XX WO200268633-A2.
XX
PD 06-SEP-2002.
XX
XX 21-NOV-2001; 2001WO-US043612.
XX
XX 22-NOV-2000; 2000US-0252500P.
XX
XX (DIAD-) DIADEXUS INC.
XX
XX Macina RA, Recipon H, Chen S, Sun Y, Liu C;
XX WPI; 2002-713376/77.
XX
XX New isolated human nucleic acid molecule and polypeptide, useful for
PT identifying, diagnosing, monitoring, staging, imaging and treating lung
PT cancer and non-cancerous diseases of the lung.
XX
XX Claim 11; Page 350-351; 389pp; English.
XX
XX The invention describes an isolated human nucleic acid (I) encoding any
CC of 120 18-1533 residue amino acid sequences (S1), given in the

CC specification, comprising any of 164 179-12421 base pair sequences (S2),
CC given in the specification. The methods and compositions of the present
CC invention are useful for identifying, diagnosing, monitoring, staging,
CC imaging and treating lung cancer and non-cancerous diseases of the lung.
CC They are also used for identifying lung tissue, monitoring and
CC identifying and/or designing antagonists of the polypeptide of the
CC invention, gene therapy, production of transgenic animals and production
CC of engineered lung tissue for treatment and research. This is the amino
CC acid sequence of a lung specific nucleic acid
XX
SQ Sequence 109 AA;
Query Match 56.2%; Score 41; DB 5; Length 109;
Best Local Similarity 41.7%; Pred. No. 1.7e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 4; Gaps 1;
QY 1 EWTN----IWW 8
DB 12 EENSEMLGVVW 23
||: |||
||: |||

RESULT 37
AAG74893
ID AAG74893 standard; protein; 121 AA.
XX
AC AAG74893;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:5657.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma.
XX Homo sapiens.
XX WO200122920-A2.
XX
PD 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US026524.
XX
XX 29-SEP-1999; 99US-0157137P.
XX
XX 03-NOV-1999; 99US-0163280P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPI; 2001-235357/24.
XX
XX N-PSDB; AAH34298.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers.
XX
XX Claim 11; Page 7203-7204; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where the
XX proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene therapy
XX and vaccine production. N and P may be used in the prevention, diagnosis
XX and treatment of diseases associated with inappropriate P expression. For
XX example, N and P may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of P by expressing inactive proteins or to
XX supplement the patient's own production of P. Additionally, N may be used
XX to produce the colon cancer-associated Ps, by inserting the nucleic acids
XX into a host cell and culturing the cell to express the proteins. N and P
XX can be used in the prevention, diagnosis and treatment of colorectal
XX carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent
XX sequences used in the exemplification of the present invention. N.B.
XX Pages 666 to 682 and page 7053 of the sequence listing were missing at

CC time of publication, meaning no sequences are present for SEQ ID NO:1027
 CC to 1052, 7921 and 7922

XX
 SQ Sequence 121 AA;
 Query Match 56.2%; Score 41; DB 4; Length 121;
 Best Local Similarity 57.1%; Pred. No. 1.9e+02;
 Matches 4; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 WTNIWW 8
 Db 17 WQQRWW 23

RESULT 38
 AAG66988
 ID AAG66988 standard; protein; 125 AA.
 AC AAG66988;
 XX
 DT 30-OCT-2001 (first entry)
 XX
 DE Human (HOMO) lamin 14 polypeptide.
 XX
 KW Human; HOMO lamin 14; cytostatic; virucidal; immunomodulatory;
 KW haemostatic; antiinflammatory; gene therapy; cancer; haemopathy;
 KW human immunodeficiency virus; HIV; infection; immunological disease;
 KW Alzheimer's disease; embryonic growth disorder.
 XX
 OS Homo sapiens.
 XX
 PN WC200164720-A1.
 XX
 PD 07-SEP-2001.
 XX
 PF 26-FEB-2001; 2001WO-CN000253.
 XX
 PR 02-MAR-2003; 2000CN-00111845.
 XX
 PA (BIOW-) BIOWINDOW GENE DEV INC SHANGHAI.
 XX
 PI Mao Y, Xie Y;
 XX
 DR WPI; 2001-550156/61.
 DR N-PSDB; AAH77641.
 XX
 PT New human lamin 14 for diagnosing and treating of Alzheimer's disease,
 PT cancer, hemopathy, human immunodeficiency virus infection, immunological
 PT diseases and phlogosis.
 XX
 PS Claim 1; Page 29; 36pp; Chinese.
 XX
 CC The invention relates to an isolated polypeptide of human (homo) lamin 14
 CC comprising a sequence of 125 amino acids or its fragment, analogue or
 CC derivative. The polypeptide and the polynucleotide encoding it are useful
 CC in the diagnosis and treatment of cancer, haemopathy, human
 CC immunodeficiency virus (HIV) infection, immunological diseases,
 CC phlogosis, Alzheimer's disease, disorders of embryonic growth, prosthetic
 CC abnormalities, neuragenesis and ecyclosis. The present sequence is the
 CC polypeptide of the invention

XX
 SQ Sequence 125 AA;
 Query Match 56.2%; Score 41; DB 4; Length 125;
 Best Local Similarity 44.4%; Pred. No. 1.9e+02;
 Matches 4; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 WTNIWW 10
 Db 70 WNAWWRR 78

RESULT 39
 AAG66987
 ID ABO000887 standard; protein; 270 AA.
 AC ABO000887;
 XX
 DT 06-AUG-2003 (first entry)
 XX
 DE Polypeptide encoded by novel human contig #138.
 XX
 KW Human; angiogenesis; cytokine; cell proliferation; pluripotent;

AA931166
 ID AA931166 standard; protein; 134 AA.
 XX
 AC AA931166;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum protein fragment SEQ ID NO: 4920.
 XX
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EP1108790-A2.
 XX
 PD 20-JUN-2001.
 XX
 PF 18-DEC-2000; 2000EP-00127688.
 XX
 PR 16-DEC-1999; 99JP-00377484.
 PR 07-APR-2000; 2000JP-00159162.
 PR 03-AUG-2000; 2000JP-00280988.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 KW Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 DR WPI; 2001-376931/40.
 DR N-PSDB; AAH66385.
 XX
 PT Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analyzing
 PT expression profile or pattern of a gene and identifying homologous gene.
 XX
 PS Claim 17; SEQ ID NO 4920; 246pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and analysing
 CC the expression profile or expression pattern of a gene derived from
 CC Coryneform bacterium, and identifying a homologue of a gene derived from
 CC coryneform bacterium. Coryneform bacteria are useful for producing amino
 CC acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described in the
 CC exemplification of the invention. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from the European Patent Office

XX
 SQ Sequence 134 AA;
 Query Match 56.2%; Score 41; DB 4; Length 134;
 Best Local Similarity 66.7%; Pred. No. 2.1e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 5 IWWNAX 10
 Db 96 LWWAR 101

RESULT 40
 ABO00887
 ID ABO00887 standard; protein; 270 AA.
 XX
 AC ABO00887;
 XX
 DT 06-AUG-2003 (first entry)
 XX
 DE Polypeptide encoded by novel human contig #138.
 XX
 KW Human; angiogenesis; cytokine; cell proliferation; pluripotent;

cell differentiation; totipotent; stem cell; transplantation; bio-sensor;
 neuroepithelial cell; autoimmune disease; neural cell; genetic disorder;
 nerve; brain tissue; central nervous system disease;
 peripheral nervous system disease; neuropathy; haematopoiesis; bone;
 myeloid disorder; lymphoid cell disorder; platelet disorder; tendon;
 regeneration; cartilage; tendon; ligament; nerve tissue growth;
 tissue repair; wound healing; burn; ulcer; osteoporosis; cancer;
 osteoarthritis; bone degenerative disorder; periodontal disease;
 gut protection; lung fibrosis; liver fibrosis; reperfusion injury;
 immune deficiency; infection; autoimmune disorder; allergic reaction;
 thrombolytic; thrombosis; coagulation disorder; hereditary disorder;
 biorhythm; circadian cycle; fertility; metabolism; catabolism; anabolism;
 neotropic; neuroprotective; antiparkinsonian; anticonvulsant;
 haemostatic; vulnery; antiulcer; osteopathic; antiarthritic;
 vasotropic; immunostimulant; antibacterial; fungicide; immunosuppressive;
 antirheumatic; antidiabetic; antiasthmatic; cytostatic; virucide.
 OS Homo sapiens.
 XX WO2003023013-A2.
 PN
 XX
 XX
 PD 20-MAR-2003.
 XX
 XX 13-SEP-2002; 2002WO-US023001.
 PF
 XX
 PR 13-SEP-2001; 2001US-0322511P.
 PR 12-SEP-2002; 2002US-00243552.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX
 XX Tang YT, Yang Y, Wang Z, Weng G, Xia Y;
 PI
 XX
 XX WPI: 2003-313249/30.
 DR N-PSDB; AC005964.
 XX
 XX Novel nucleic acids and polypeptides for diagnosis, treatment of central
 PT and peripheral nervous system diseases and neuropathies, such as
 PT Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 PT lateral sclerosis.
 XX
 PS Example 3; SEQ ID NO 1011; 300pp; English.
 XX
 CC The present invention relates to the isolation of novel human
 CC polynucleotide sequences and their encoding polypeptides. The novel
 CC polypeptides exhibit activities relating to angiogenesis, cytokine, cell
 CC proliferation, cell differentiation, antiinflammatory, and stem cell
 CC growth factor activities. The polypeptides are involved in the
 CC proliferation, differentiation and survival of pluripotent and totipotent
 CC stem cells, and are useful for re-engineering damaged or diseased
 CC tissues, transplantation, manufacture of bio-pharmaceuticals and
 CC development of bio-sensors. The polypeptides can be used to manipulate
 CC stem cells in culture to give rise to neuroepithelial cells that can be
 CC used to augment or replace cells damaged by illness, autoimmune disease,
 CC accidental damage or genetic disorders. The polypeptides induce the
 CC proliferation of neural cells and regeneration of nerve and brain tissue
 CC and are useful for the treatment of central and peripheral nervous system
 CC diseases and neuropathies, such as Alzheimer's, Parkinson's disease,
 CC Huntington's disease, amyotrophic lateral sclerosis (ALS). The
 CC polypeptides are also involved in chemotactic or chemokinetic activity,
 CC regulation of haematopoiesis and are useful for treating myeloid or
 CC lymphoid cell disorders, platelet disorders such as thrombocytopaenia and
 CC for regeneration of bone, cartilage, tendon, ligament and/or nerve tissue
 CC growth, in tissue repair, healing of burns, incisions, ulcers, for
 CC treating osteoporosis, osteoarthritis, bone degenerative disorders, and
 CC periodontal disease. The polypeptides are also useful for gut protection
 CC or regeneration and treatment of lung or liver fibrosis, reperfusion
 CC injury in various tissues, various immune deficiencies and disorders
 CC including severe combined immunodeficiency (SCID), bacterial or fungal
 CC infections, autoimmune disorders (e.g. multiple sclerosis, rheumatoid
 CC arthritis, diabetes mellitus, myasthenia gravis), allergic reactions and
 CC conditions, such as asthma or other respiratory problems. The
 CC polypeptides are involved in thrombolysis or thrombosis and are useful in
 CC treatment of various coagulation disorders (including hereditary

CC disorders such as haemophilia) or to enhance coagulation and other
 CC haemostatic events in treating wounds resulting from trauma, surgery or
 CC other causes. The polypeptides exhibit immune stimulating or immune
 CC suppressing activity, and are useful for treating autoimmune diseases or
 CC cancer. They also inhibit the growth, infection or function of infectious
 CC agents such as bacteria, fungi, viruses, effect biorhythms or circadian
 CC cycles of rhythms, fertility of male or female subjects, metabolism,
 CC catabolism, and anabolism. ABO00750-ABO00950 represent polypeptides
 CC encoded by novel contigs assembled in the examples of the present
 CC invention. Note: the sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 270 AA;
 Query Match 56.2%; Score 41; DB 6; Length 270;
 Best Local Similarity 71.4%; Pred. No. 4.1e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EWTNIWW 7
 |||:
 Db 196 EWTSDWW 202
 RESULT 41
 ABO14998
 ID ABO14998 standard; protein; 344 AA.
 XX
 AC ABO14998;
 XX
 DT 26-AUG-2003 (first entry)
 XX
 DE Human MCV11 protein.
 XX
 KW Human; MOVX; inflammatory disorder; demyelination disease; stroke;
 KW renal disorder; infection; cardiomyopathy; atherosclerosis; acne;
 KW hypertension; pancreatitis; Von Hippel-Lindau; endometriosis; fertility;
 KW scleroderma; cirrhosis; inflammatory bowel disease; Crohn's disease;
 KW haemophilia; autoimmune disease; allergy; AIDS;
 KW graft versus host disease; Alzheimer's disease; arthritis; pain;
 KW Parkinson's disease; Huntington's disease; obesity; diabetes;
 KW hair growth; hair loss; asthma; schizophrenia; glomerulonephritis;
 KW lupus erythematosus; psoriasis; antidiabetic; anorectic; metabolic;
 KW neotropic; neuroprotective; cytostatic; antibacterial; virucide;
 KW protozoacide; antiarteriosclerotic; hypotensive; cerebroprotective;
 KW antiinflammatory; gynaecological; antinfertility; dermatological;
 KW hepatotropic; haemostatic; immunosuppressive; antiasthmatic;
 KW antiarthritic; anticonvulsant; antiseborrhoeic; antiallergic;
 KW neuroleptic; anti-HIV; analgesic; nephrotropic; antipsoriatic.
 XX
 OS Homo sapiens.
 XX
 XX WO200298917-A2.
 XX
 PD 12-DEC-2002.
 XX
 XX 12-FEB-2002; 2002WO-US022049.
 XX
 PR 12-FEB-2001; 2001US-0268221P.
 PR 13-FEB-2001; 2001US-0268496P.
 PR 14-FEB-2001; 2001US-0268646P.
 PR 14-FEB-2001; 2001US-0268665P.
 PR 15-FEB-2001; 2001US-0269136P.
 PR 16-FEB-2001; 2001US-0269310P.
 PR 16-FEB-2001; 2001US-0269530P.
 PR 16-MAR-2001; 2001US-0276405P.
 PR 16-MAR-2001; 2001US-0276399P.
 PR 16-MAR-2001; 2001US-0276703P.
 PR 23-MAR-2001; 2001US-0278199P.
 PR 28-MAR-2001; 2001US-0279274P.
 PR 30-MAR-2001; 2001US-0280238P.
 PR 02-APR-2001; 2001US-0280899P.
 PR 08-AUG-2001; 2001US-0310797P.

OS Homo sapiens.
 XX Key
 FT Peptide
 FT 1. .34 Location/Qualifiers
 FT /label= Signal_peptide
 FT 7. .12
 FT Modified-site
 FT /note= "N-myrisotylated"
 FT 30. .35
 FT Modified-site
 FT /note= "N-myrisotylated"
 FT 35. .374
 FT Protein
 FT /label= Mature_protein
 FT 38. .43
 FT Modified-site
 FT /note= "N-myrisotylated"
 FT 40. .50
 FT Region
 FT /note= "prokaryotic membrane lipoprotein lipid attachment site"
 FT 44. .49
 FT Modified-site
 FT /note= "N-myrisotylated"
 FT 59. .64
 FT Modified-site
 FT /note= "N-myrisotylated"
 FT 117. .120
 FT Modified-site
 FT /note= "CAMP- and cGMP-dependent protein kinase phosphorylation site"
 FT 118. .125
 FT Modified-site
 FT /note= "tyrosine kinase phosphorylation site"
 FT 132. .135
 FT Modified-site
 FT /note= "casein kinase II phosphorylation site"
 FT 141. .144
 FT Modified-site
 FT /note= "Asn is N-glycosylated"
 FT 147. .150
 FT Modified-site
 FT /note= "casein kinase II phosphorylation site"
 FT 210. .213
 FT Modified-site
 FT /note= "casein kinase II phosphorylation site"
 FT 221. .224
 FT Modified-site
 FT /note= "Amidation"
 FT 227. .230
 FT Modified-site
 FT /note= "casein kinase II phosphorylation site"
 FT 234. .236
 FT Modified-site
 FT /note= "protein kinase C phosphorylation site"
 FT 237. .242
 FT Modified-site
 FT /note= "N-myrisotylated"
 FT 254. .259
 FT Modified-site
 FT /note= "N-myrisotylated"
 FT 270. .273
 FT Modified-site
 FT /note= "casein kinase II phosphorylation site"
 FT 311. .314
 FT Modified-site
 FT /note= "casein kinase II phosphorylation site"
 FT 323. .326
 FT Modified-site
 FT /note= "casein kinase II phosphorylation site"
 FT 323. .325
 FT Modified-site
 FT /note= "protein kinase C phosphorylation site"
 FT 332. .337
 FT Modified-site
 FT /note= "N-myrisotylated"
 FT
 XX WO200050443-A2.
 PN
 XX
 PD 31-AUG-2000.
 XX
 PF 25-FEB-2000; 2003WO-US005035.
 XX
 PR 26-FEB-1999; 99US-00259387.
 XX
 XX (MILL-) MILLENNIUM PHARM INC.
 PA
 XX Fraser CC;
 PT
 XX WPI; 2000-533178/48.
 DR
 DR N-PSDB; AAA50442.
 XX
 PT Nucleic acids encoding TANGO 228, 240 and 243 pp. which have homology to
 PT the rat mast cell Ag-32, the Mycobacterium tuberculosis hypothetical
 PT protein Rv0712 and human phospholipase A2-activating protein.
 XX
 XX

PS
 XX Claim 3; Fig 6A-B; 188pp; English.
 CC The present sequence is that of human TANGO 240, a 40.6 kDa secreted
 CC protein with homology to the Mycobacterium tuberculosis hypothetical
 CC protein Rv0712. The sequence was deduced from that of a cDNA clone
 CC isolated from an osteoblast cDNA library. TANGO 240 proteins, nucleic
 CC acids, and modulators can be used to: modulate the proliferation,
 CC differentiation and/or function of bone and cartilage cells, e.g.
 CC chondrocytes and osteoblasts, and to treat bone and/or cartilage
 CC associated diseases and disorders, such as those resulting from injury
 CC (bone breakage, cartilage tearing), degeneration (osteoporosis), and
 CC degeneration of joints (arthritis, osteoarthritis, bone wearing).
 CC Homology to Rv0712 suggests use to treat diseases associated with
 CC bacterial infection, e.g. tuberculosis. TANGO 240 can also be used to
 CC modulate (e.g. trigger) an immune response to treat an immunological
 CC disease such as those associated with the respiratory system, e.g.
 CC asthma. TANGO 240 proteins, nucleic acids and modulates can also be used
 CC to treat disorders of the cells and tissues in which TANGO 240 is
 CC expressed i.e. heart, brain, spleen, lung, liver, skeletal muscle,
 CC kidney, testis, colon, thymus, peripheral blood leukocytes, small
 CC intestine and placenta, e.g. ischaemic heart disease, atherosclerosis,
 CC head trauma, brain cancer, splenic lymphoma, splenomegaly, lung cancer,
 CC cystic fibrosis, rheumatoid lung disease, liver cirrhosis, hepatitis,
 CC muscular dystrophy, stroke, muscular atrophy, glomerulonephritis, end
 CC stage renal disease, uraemia, testicular cancer, colon cancer and colonic
 CC volvulus, DiGeorge syndrome, thymoma, autoimmune disorders, atresia,
 CC Crohn's disease, and placental disorders. TANGO 240 polypeptides can be
 CC obtained by recombinant DNA methods or expressed using gene therapy
 CC methods. They can also be used to raise antibodies (useful for diagnosis),
 CC and to screen for modulator compounds
 XX
 SQ Sequence 374 AA;
 Query Match 56.2%; Score 41; DB 3; Length 374;
 Best Local Similarity 71.4%; Pred. No. 5.7e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EWTNIWN 7
 |||:
 Db 300 EWTSDWN 306
 RESULT 44
 AAB87396
 ID AAB87396 standard; protein; 374 AA.
 XX
 AC AAB87396;
 XX
 DT 22-MAY-2001 (first entry)
 XX
 DE Human gene 8 encoded secreted protein HMAV121, SEQ ID NO:137.
 XX
 KW Human; secreted protein; proliferative disorder; cancer; tumour;
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;
 KW gastrointestinal disorder; pregnancy-related disorder;
 KW endocrine disorder; infection; wound healing; vulnerability; cell culture;
 KW chemotaxis; food additive; binding partner identification.
 XX
 OS Homo sapiens.
 XX
 PN WO200118022-A1.
 XX
 PD 15-MAR-2001.
 XX
 PF 31-AUG-2000; 2000WO-US024008.
 XX
 PR 03-SEP-1999; 99US-0152315P.
 PR 03-SEP-1999; 99US-0152317P.

XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX
PI Ni J, Baker KP, Birse CE, Fiscella M, Komatsoulis GA, Rosen CA;
PI Soppet DR, Young PE, Ebner R, Duan DR, Olsen HS, Lafleur DW;
PI Moore PA, Shi Y, Wei Y, Florence KA;
XX
XX WPI: 2001-203081/20.
DR N-PSDB; AAF91912.
XX
XX Nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers.
XX
XX Claim 11; Page 565-566; 607pp; English.
XX
XX AAF91858-AAF91929 represent cDNAs corresponding to 52 human secreted
CC protein genes, and AAB87342-AAB87413 represent the proteins they encode.
CC The genes AAB87414-AAB87454 represent human secreted protein fragments. The genes
CC and their corresponding secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 52 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin
CC disorders (e.g. psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiotensin-related disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunoassay (ELISA). The present sequence represents a human
CC secreted protein of the invention.
XX
XX Sequence 374 AA;
SQ
Query Match 56.2%; Score 41; DB 4; Length 374;
Best Local Similarity 71.4%; Pred. No. 5.7e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 EWTNIW 7
Db 300 EWTSDW 306
RESULT 45
AAE12629
ID AAE12629 standard; protein; 374 AA.
XX
XX AAE12629;
AC
XX
XX 03-JAN-2002 (first entry)
XX
XX Human gene 7 encoded secreted protein HMA121.
XX
XX Human; ovarian cancer antigen; proliferative disorder; cancer; tumour;
KW respiratory system disorder; asthma; haematopoietic disorder; skin aging;
KW immune system disorder; AIDS; skin disorder; autoimmune disease; allergy;
KW rheumatoid arthritis; inflammation; neurological disorder; septic shock;
KW Alzheimer's disease; Parkinson's disease; diabetes; angiogenic disorder;
KW atherosclerosis; cardiovascular disorder; gastrointestinal disorder;
XX

KW epithelial cell proliferation; transplantation; chemotaxis; infection;
KW food additive; wound healing; endocrine disorder; kidney disorder;
XX gene therapy; cytostatic; chromosome 18p11.2.
XX Homo sapiens.
XX WO200170804-A1.
XX 27-SEP-2001.
XX
XX 16-MAR-2001; 2001WO-US008585.
XX
XX 17-MAR-2000; 2000US-0190076P.
XX 23-AUG-2000; 2000US-0227009P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Birse CE, Rosen CA;
XX WPI: 2001-639119/73.
XX N-PSDB; RAD20624.
XX
XX Nucleic acids encoding 7 ovarian cancer antigen polypeptides, useful for
PT preventing, diagnosing and/or treating ovarian cancer and ovarian cancer
PT metastases.
XX
XX Claim 11; Page 417-418; 427pp; English.
XX
XX AAD20618-AAD20624 represent cDNAs corresponding to 7 human ovarian and
CC ovarian cancer associated protein (collectively known as ovarian cancer
CC antigens) genes, and AAB12623-AAE12629 represent the proteins they
CC encode. AAE12630-AAE12638 represent human ovarian cancer antigenic
CC fragments or variants. Ovarian cancer antigens and their corresponding
CC DNAs are used in the prevention, diagnosis and treatment of diseases
CC associated with their inappropriate expression. These disorders include
CC proliferative disorders, cancer, tumours, respiratory system disorders,
CC asthma, haematopoietic disorders, diseases of the immune system, AIDS,
CC skin disorders, autoimmune diseases (e.g., rheumatoid arthritis),
CC inflammation, allergic, neurological disorders (e.g., Alzheimer's
CC disease, Parkinson's disease), septic shock, diabetes, atherosclerosis,
CC cardiovascular disorders, angiotensin disorders, kidney disorders,
CC gastrointestinal disorders, endocrine disorders and infections. The
CC proteins can also be used to aid wound healing and epithelial cell
CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, to identify their cognate ligands or binding
CC partners, in chemotaxis and can be used as a food additive. Antibodies
CC specific for a protein of the invention can be used in alleviating
CC symptoms associated with the disorders mentioned above and in diagnostic
CC immunoassays e.g., enzyme linked immunosorbent assay (ELISA). The DNA of
CC the invention is used in gene therapy. The present sequence represents a
CC human ovarian cancer antigen of the invention
XX
XX Sequence 374 AA;
SQ
Query Match 56.2%; Score 41; DB 4; Length 374;
Best Local Similarity 71.4%; Pred. No. 5.7e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 EWTNIW 7
Db 300 EWTSDW 306
RESULT 46
AAB88402
ID AAB88402 standard; protein; 374 AA.
XX
XX AAB88402;
AC
XX 23-MAY-2001 (first entry)
XX
XX Human membrane or secretory protein clone P5EC0152.
DE
XX

KW Human; secretory protein; membrane protein; vaccine; gene therapy;
 KW rheumatoid arthritis; diabetes.
 XX Homo sapiens.
 OS
 XX
 PN EP1067182-A2.
 XX
 XX
 PD 10-JAN-2001.
 XX
 XX
 PF 07-JUL-2000; 2000EP-00114090.
 XX
 XX
 PR 08-JUL-1999; 99JP-00194179.
 PR 11-JAN-2000; 2000JP-00118775.
 PR 02-MAY-2000; 2000JP-00183766.
 XX
 XX (HELI-) HELIX RES INST.
 XX
 XX Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;
 XX WPI; 2001-093989/11.
 DR N-PSDB; AAF93829.
 XX
 XX Nucleic acids encoding secretory proteins/membrane proteins, useful in
 PT gene therapy or as candidate target molecules in drug development.
 XX
 XX Claim 1; SEQ ID NO 172; 609pp + Sequence Listing; English.
 XX
 CC This invention relates to nucleic acid sequences AAF93744 - AAF93916
 CC which encode human secretory or membrane proteins represented by AAB88317
 CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and
 CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the
 CC invention. The invention also includes methods for the production of
 CC antibodies directed against the proteins, and cDNA sequences, which can
 CC be used in vaccines. The polynucleotide sequences can be used in gene
 CC therapy. The polynucleotide sequences and the proteins they encode may be
 CC used in the prevention, treatment and diagnosis of diseases associated
 CC with inappropriate secretory protein/membrane protein expression. The
 CC nucleic acids and complementary sequences may also be used as DNA probes
 CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect
 CC and quantitate the presence of similar nucleic acid sequences in samples.
 CC They may also be used to study the expression and function of secretory
 CC proteins/membrane polypeptides and their role in metabolism. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC against them and in assays to identify modulators (agonists and
 CC antagonists) of expression and activity. The antibodies and antagonists
 CC may also be used as therapeutic agents to down regulate expression and
 CC activity. The antibodies may also be used as diagnostic agents for
 CC detecting the presence of the polypeptides in samples (e.g. by enzyme
 CC linked immunosorbent assay (ELISA)). Examples of diseases which may be
 CC treated include rheumatoid arthritis and diabetes
 XX
 SQ Sequence 374 AA;
 Query Match 56.2%; Score 41; DB 4; Length 374;
 Best Local Similarity 71.4%; Pred. No. 5.7e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 EWTNIWW 7
 Db 300 EWTSDWW 306
 RESULT 47
 AAB85784
 ID AAB85784 standard; protein; 374 AA.
 XX
 AC AAB85784;
 XX
 DT 29-OCT-2001 (first entry)
 XX
 DE Human kinase PKIN-3.
 XX
 XX PKIN; kinase; cytostatic; immunosuppressive; immunostimulant; human;
 KW
 KW Human; secretory protein; membrane protein; vaccine; gene therapy;
 KW antiarteriosclerotic; cardiant; gene therapy; antisense therapy.
 XX Homo sapiens.
 OS
 XX
 PN WO200150991-A2.
 XX
 XX
 PD 23-AUG-2001.
 XX
 XX
 PF 16-FEB-2001; 2001WO-US005240.
 XX
 XX
 PR 17-FEB-2000; 2000US-0183682P.
 PR 02-MAR-2000; 2000US-0186559P.
 PR 09-MAR-2000; 2000US-0188606P.
 PR 17-MAR-2000; 2000US-0189998P.
 PR 30-MAR-2000; 2000US-0193851P.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 XX
 XX Tang YT, Buford N, Gandhi AR, Patterson C, Khan FA, Yue H;
 XX Hafalia A, Shih LL, Tribouley CM, Yao MG, Burrill JD, Marcus GA;
 PI Zingler KA, Lu DAM, Bandman O, Policky JL, Griffin JA, Thornton M;
 PI Nguyen DB, Lal P, Walsh RT;
 XX
 XX WPI; 2001-514771/56.
 DR N-PSDB; AAF76211.
 XX
 XX isolated human kinase polypeptides useful in the diagnosis, treatment and
 PT prevention of cancer, immune disorders and disorders affecting growth and
 PT development.
 XX
 XX Claim 1; Page 105; 126pp; English.
 XX
 CC The invention provides human kinases (PKIN) and polynucleotides encoding
 CC PKIN. The PKIN polypeptides can be expressed using standard recombinant
 CC methodology. The PKIN polypeptides, polynucleotides, modulators and
 CC specific antibodies are useful in the diagnosis, treatment and prevention
 CC of cancer, immune disorders, disorders affecting growth and development,
 CC atherosclerosis, and other cardiovascular diseases, and lipid disorders
 CC and in the assessment of the effects of exogenous compounds on the
 CC expression of nucleic acid sequences of human kinases. The present
 CC sequence represents a human PKIN-3 polypeptide
 XX
 SQ Sequence 374 AA;
 Query Match 56.2%; Score 41; DB 4; Length 374;
 Best Local Similarity 71.4%; Pred. No. 5.7e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 EWTNIWW 7
 Db 300 EWTSDWW 306
 RESULT 48
 AAB65345
 ID AAB65345 standard; protein; 374 AA.
 XX
 AC AAB65345;
 XX
 DT 27-AUG-2002 (first entry)
 XX
 DE Human albumin fusion protein #2020.
 XX
 XX Albumin fusion protein; therapeutic protein X; human albumin; HA;
 KW human serum albumin; HSA; cancer; reproductive disorder;
 KW digestive disorder; immune disorder; endocrine disorder;
 KW haematopoietic disorder; neural disorder; connective disorder;
 KW cytostatic; antiinfectivity; antiinflammatory; antitumor;
 KW immunomodulator; anti-HIV; antidiabetic; haemostatic; neurotropic;
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
 KW osteopathic; antiarthritic.
 XX
 OS Homo sapiens.

PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian
PT cancer), immune disorders, cardiovascular disorders and neurological
XX diseases.

PS Claim 11; SEQ ID NO 2819; 2922pp; English.

XX The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and
CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vaginitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents a human ovarian antigen of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from: WIPO
CC at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 403 AA;

Query Match 56.2%; Score 41; DB 5; Length 403;
Best Local Similarity 71.4%; Pred. No. 6.1e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNIWV 7
|||: ||
Db 329 EWTSDWV 335

RESULT 51
AAV02610
ID AAY02610 standard; protein; 416 AA.

XX AC AAY02610;

XX DT 21-JUL-1999 (first entry)

XX Amino acid sequence of a putative efflux protein.

XX Putative efflux protein; Pseudomonas mendocina KR-1; resistance; solvent;
KW antibiotic; p-hydroxybenzoic acid.

XX Pseudomonas mendocina.

XX WQ9923224-A.

XX PD 14-MAY-1999.

XX PF 30-OCT-1998; 98WO-US023266.

XX PR 31-OCT-1997; 97US-00961738.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Chen KK;

XX WPI; 1999-313345/26.
DR N-PSDB; AAX59540.

XX New putative efflux protein gene increases resistance to solvents and
PT antibiotics.

XX Claim 4; Page 32-33; 28pp; English.

XX The present sequence represents a putative efflux protein obtained from
CC Pseudomonas mendocina KR-1. The protein increases resistance to solvents
CC and antibiotics, and is used particularly for constructing strains that
CC produce p-hydroxybenzoic acid. Overexpressing the efflux system or its
CC expression from a plasmid vector increases resistance of bacteria to a
CC variety of toxic substances, while inactivating an efflux system
CC increases sensitivity to antibiotics and toxic substances
XX Sequence 416 AA;

Query Match 56.2%; Score 41; DB 2; Length 416;
Best Local Similarity 57.1%; Pred. No. 6.3e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNIWV 7
|||: ||
Db 220 EWRSVWV 226

RESULT 52
AAB98023
ID AAB98023 standard; protein; 416 AA.

XX AC AAB98023;

XX DT 16-AUG-2001 (first entry)

XX Pseudomonas mendocina KR-1 putative efflux 4.5X protein SEQ ID NO:2.

XX Pseudomonas mendocina KR-1; efflux; 4.5X; RFLP marker; genetic mapping;
KW restriction fragment length polymorphism marker; physical mapping.
XX solvent tolerance.

XX Pseudomonas mendocina.

XX US6225083-B1.

XX PD 01-MAY-2001.

XX PF 15-JUN-1999; 99US-00333208.

XX PR 31-OCT-1997; 97US-00961738.

XX PR 30-OCT-1998; 98US-00181270.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Chen KK;

XX WPI; 2001-342396/36.

DR N-PSDB; AAH21716.

XX Novel isolated Pseudomonas efflux polynucleotide, useful for probing an
PT organism's efflux system to gain understanding of the mechanism of
PT solvent tolerance and as restriction fragment length polymorphism marker.

XX Claim 1; Col 23-26; 17pp; English.

XX The present sequence represents a putative efflux protein, designated
CC 4.5X, isolated from Pseudomonas mendocina KR-1. The putative efflux 4.5X
CC gene (I) is located immediately downstream of the poba gene in
CC Pseudomonas mendocina KR-1. The present invention also describes: (1) a
CC chimeric gene (II) comprising (i) operably linked to at least one
CC suitable regulatory sequence; (2) a transformed host cell (III)
CC comprising (II); (3) isolation of (I); and (4) a complement of (I). (II)

CC is useful for altering the level of expression of pseudomonas efflux
 CC protein in a host cell by transforming the cell with (II), and growing
 CC the transformed host cell produced under conditions that are suitable for
 CC expression of (III), resulting in production of increased levels of
 CC Pseudomonas efflux proteins in the transformed host cell relative to
 CC expression levels of an untransformed host cell. (I) is useful for
 CC isolating cDNAs and genes encoding a homologous putative efflux 4.5x gene
 CC from the same or other bacterium species. (I) is useful as a restriction
 CC fragment length polymorphism (RFLP) marker, and for carrying out nucleic
 CC acid amplification-based methods of genetic and physical mapping. (I) is
 CC useful for probing an organism's efflux system to gain an understanding
 CC of the mechanism of solvent tolerance

XX SQ Sequence 416 AA;

Query Match 56.2%; Score 41; DB 4; Length 416;
 Best Local Similarity 57.1%; Pred. No. 6.3e+02;
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EWTNIWW 7
 || :||
 Db 220 EWRSVWW 226

RESULT 53

AB84207
 ID AAB84207 standard; protein; 416 AA.

XX AC AAB84207;

XX DT 06-AUG-2001 (first entry)

XX DE Amino acid sequence of a putative efflux gene.

XX KW Efflux 4.5x protein; Pseudomonas mendocina KR-1; toxic compound.

XX OS Pseudomonas mendocina.

XX PN US6235882-B1.

XX PD 22-MAY-2001.

XX PF 15-JUN-1999; 99US-00333254.

XX PR 31-OCT-1997; 97US-00961738.

XX PR 30-OCT-1998; 98US-00183270.

XX PA (DUPO) DU PONT DE NEMOURS & CO E I.

XX PI Chen KK;

XX DR WPI; 2001-366425/38.

XX DR N-PSDB; AAF90209.

PT New gene encoding a putative efflux protein for solvents or antibiotics
 PT isolated from Pseudomonas mendocina, useful for isolating cDNAs and genes
 PT encoding a homologous putative efflux 4.5x gene from other bacterial
 PT species.

XX FS Claim 2; Col 21-24; 16pp; English.

XX The present sequence represents a putative efflux 4.5x protein. It is
 CC isolated from Pseudomonas mendocina KR-1. Efflux proteins play an
 CC important role in bacteria by conferring resistance to toxic compounds
 CC such as aromatic solvents. The efflux 4.5x nucleic acid may be used to
 CC isolate cDNAs and genes encoding a homologous putative efflux 4.5x gene
 CC from the same or other bacterial species. The nucleic acids may also be
 CC used as restriction fragment length polymorphism marker, as probes, for
 CC physical mapping placement sequences on physical maps, in direct
 CC fluorescence in situ hybridization mapping, and in various nucleic acid-
 CC based methods of genetic and physical mapping. The 4.5x peptides may be
 CC used to immunize animals to produce monoclonal or polyclonal antibodies
 CC specific for these peptides

XX SQ Sequence 416 AA;

Query Match 56.2%; Score 41; DB 4; Length 416;
 Best Local Similarity 57.1%; Pred. No. 6.3e+02;
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EWTNIWW 7
 || :||
 Db 220 EWRSVWW 226

RESULT 54

AB891683
 ID AB891683 standard; protein; 692 AA.

XX AC AB891683;

XX DT 31-MAY-2002 (first entry)

XX DE Herbicidally active polypeptide SEQ ID NO 894.

XX KW Herbicidal; plant; agriculture; herbicide.

XX OS Arabidopsis thaliana.

XX PN WO200210210-A2.

XX PD 07-FEB-2002.

XX PF 28-AUG-2001; 2001WO-EP009892.

XX PR 28-AUG-2001; 2001WO-EP009892.

XX PA (FARB) BAYER AG.

XX PI Tietjen K, Weidier M;

XX DR WPI; 2002-269010/31.

XX Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms.

XX PS Claim 5; SEQ ID NO 894; 261pp + Sequence Listing; English.

XX The invention relates to identifying target proteins (ABB90790-ABB94016)
 CC for herbicidally active compounds, comprising aligning and comparing
 CC nucleic acid or amino acid sequences from plant with nucleic acid or
 CC amino acid sequences from non-plant organisms using suitable search
 CC parameters, where plant sequences having an E-value greater by a factor
 CC of 3 than the E-value of most similar non-plant sequences are selected.
 CC The polypeptides or nucleic acids encoding them are useful for
 CC identifying modulators. The identified modulators are useful as
 CC herbicides

XX SQ Sequence 692 AA;

Query Match 56.2%; Score 41; DB 5; Length 692;
 Best Local Similarity 50.0%; Pred. No. 1e+03;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EWTNIWW 8
 :|| :||
 Db 345 QWGRDWW 352

RESULT 55

ABG03509

ID ABG03509 standard; protein; 732 AA.

XX AC ABG03509;

XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #3500.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Dmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS67696.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX PS Claim 20; SEQ ID NO 33668; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 732 AA;

Query Match 55.2%; Score 41; DB 4; Length 732;
Best Local Similarity 57.1%; Pred. No. 1.1e+03;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 WTNIWWW 8
DB 718 WQSWWW 724

RESULT 56
ABB91045
ID ABB91045 standard; protein; 764 AA.
XX AC ABB91045;
XX DT 31-MAY-2002 (first entry)

XX Herbicidally active polypeptide SEQ ID NO 256.
XX DE Herbicidal; plant; agriculture; herbicide.
XX KW Arabidopsis thaliana.
XX OS WO200210210-A2.
XX PN 07-FEB-2002.
XX PD 28-AUG-2001; 2001WO-EP009892.
XX PF 28-AUG-2001; 2001WO-EP009892.
XX PR (FARB) BAYER AG.
XX PA Tietjen K, Weidler M;
XX PI WPI; 2002-269010/31.
XX DR Identifying plant target proteins for herbicidally active compounds,
XX PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms.
XX PS Claim 5; SEQ ID NO 256; 261pp + Sequence Listing; English.
XX CC The invention relates to identifying target proteins (ABB90790-ABB94016)
CC for herbicidally active compounds, comprising aligning and comparing
CC nucleic acid or amino acid sequences from plant with nucleic acid or
CC amino acid sequences from non-plant organisms using suitable search
CC parameters, where plant sequences having an E-value greater by a factor
CC of 3 than the E-value of most similar non-plant sequences are selected.
CC The polypeptides or nucleic acids encoding them are useful for
CC identifying modulators. The identified modulators are useful as
XX herbicides
XX SQ Sequence 764 AA;

Query Match 56.2%; Score 41; DB 5; Length 764;
Best Local Similarity 83.3%; Pred. No. 1.2e+03;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 IWWWAK 10
DB 397 IWWWRK 402

RESULT 57
ABG17639
ID ABG17639 standard; protein; 936 AA.
XX AC ABG17639;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #17630.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX DT

DR N-PSDB; AA192169.
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
PS Claim 20; SEQ ID NO 26130; 1399pp + Sequence listing; English.
XX The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 40 AA;

Query Match 54.8%; Score 40; DB 4; Length 40;
Best Local Similarity 57.1%; Pred. No. 85;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 TNIIWWWA 9
Db 20 SQLWWWA 26
: : : : :
: : : : :

RESULT 60
ABP99836
ID ABP99836 standard; protein; 57 AA.

XX AC ABP99836;
XX DT 26-MAR-2003 (first entry)
XX DE Human secreted protein SEQ ID NO 780.
XX Human; secreted protein; neurotropic; neuroprotective; cytostatic;
KW virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW vulnary; antibacterial; antiparkinsonian; antiskickling; antianaemic;
KW antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;
KW antiinflammatory; antiallergic; antidiabetic; antitumor; anticonvulsant;
KW antifungal; antiparasitic; cardiant; immune disorder; infection; vaccine;
KW cardiovascular disorder; neurological disease; nephrotropic;
KW gene therapy.
XX OS Homo sapiens.
XX FN WO200277186-A2.
XX DT 03-OCT-2002.
XX PF 26-MAR-2002; 2002WO-US009188.
XX PR 27-MAR-2001; 2001US-0278650P.
XX PR 12-SEP-2001; 2001US-00950082.
XX PR 12-SEP-2001; 2001US-00950083.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX WPI; 2003-040583/03.
XX DR N-PSDB; ABZ67257.
XX New human secreted proteins encoded by genes contained in cDNA clones
PT (e.g. HGAC19), useful for preventing, treating or diagnosing e.g. AIDS,
PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or

PT West Nile fever.
XX Claim 1; Page 1497-1498; 2423pp; English.
XX The invention relates to novel human genes (ABZ66891-ABZ68209) and the
CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,
CC treating or ameliorating medical conditions e.g. by protein or gene
CC therapy. The genes are isolated from a range of human tissues disclosed
CC in the specification. The nucleic acids, proteins, antibodies and
CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
CC bacterial, fungal and parasitic infections
XX
SQ Sequence 57 AA;

Query Match 54.8%; Score 40; DB 6; Length 57;
Best Local Similarity 57.1%; Pred. No. 1.2e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 TNIIWWWA 9
Db 45 SHIWWWS 51
: : : : :
: : : : :

RESULT 61
ABP99855
ID ABP99855 standard; protein; 57 AA.

XX AC ABP99855;
XX DT 26-MAR-2003 (first entry)
XX DE Human secreted protein SEQ ID NO 799.
XX Human; secreted protein; neurotropic; neuroprotective; cytostatic;
KW virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW vulnary; antibacterial; antiparkinsonian; antiskickling; antianaemic;
KW antiarthritic; cancer; antirheumatic; hepatotropic; cerebroprotective;
KW antiinflammatory; antiallergic; antidiabetic; antitumor; anticonvulsant;
KW antifungal; antiparasitic; cardiant; immune disorder; infection; vaccine;
KW cardiovascular disorder; neurological disease; nephrotropic;
KW gene therapy.
XX OS Homo sapiens.
XX FN WO200277186-A2.
XX DT 03-OCT-2002.
XX PF 26-MAR-2002; 2002WO-US009188.
XX PR 27-MAR-2001; 2001US-0278650P.
XX PR 12-SEP-2001; 2001US-00950082.
XX PR 12-SEP-2001; 2001US-00950083.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM;
XX WPI; 2003-040583/03.
XX DR N-PSDB; ABZ67276.
XX New human secreted proteins encoded by genes contained in cDNA clones
PT (e.g. HGAC19), useful for preventing, treating or diagnosing e.g. AIDS,
PT multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or
PT West Nile fever.

XX Claim 1; Page 1505; 2423pp; English.

XX The invention relates to novel human genes (ABZ66891-ABZ68209) and the
 CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,
 CC treating or ameliorating medical conditions e.g. by protein or gene
 CC therapy. The genes are isolated from a range of human tissues disclosed
 CC in the specification. The nucleic acids, proteins, antibodies and
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
 CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
 CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
 CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.
 CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
 CC bacterial, fungal and parasitic infections

XX SQ Sequence 57 AA;
 Query Match 54.8%; Score 40; DB 6; Length 57;
 Best Local Similarity 57.1%; Pred. No. 1.2e+02;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 3 TNIWWWA 9
 Db 45 SHIWWWS 51
 ::|||::

RESULT 62
 ABR01328
 ID ABR01328 standard; protein; 57 AA.
 AC ABR01328;
 XX 12-MAY-2003 (first entry)
 DT Human gene 382-encoded secreted protein HLMAU42, SEQ ID NO:809.
 DE Human; secreted protein; cancer; tumour; hyperproliferative disorder;
 XX autoimmune disorder; inflammation; angiogenic diseases; AIDS;
 KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
 KW drug screening; chromosome identification; chromosome mapping;
 KW cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
 KW antianaemic; vulnery.
 XX Homo sapiens.
 OS WO200277013-A2.
 PK 03-OCT-2002.
 XX 26-MAR-2002; 2002WO-US009370.
 XX 27-MAR-2001; 2001US-0278650P.
 PR 12-SEP-2001; 2001US-00950082.
 PR 12-SEP-2001; 2001US-00950083.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Ruben SM;
 PI WPI; 2003-040578/03.
 XX N-PSDB; ABZ73662.
 DR New human secreted proteins and nucleic acids, useful for detecting or
 PT treating cancer or other hyperproliferative disorders, autoimmune
 PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia.
 XX Claim 13; Page 1498; 2474pp; English.
 PS ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted

CC protein genes, and ABP00947-ABP01363 represent the proteins they encode.
 CC ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The
 CC invention also encompasses antibodies specific for the secreted proteins,
 CC the use of the secreted proteins in drug screening and recombinant
 CC vectors and host cells comprising a nucleic acid of the invention. The
 CC secreted proteins are thought to be involved in biological activities
 CC associated with cellular signalling, cellular differentiation, cell
 CC migration, prohormone activation and neurotransmitter activity. The
 CC secreted proteins, nucleic acids encoding them, antibodies or antibody
 CC fragments specific for the secreted proteins, and modulators of protein
 CC activity are useful for diagnosing or treating cancers or other
 CC hyperproliferative disorders. Additionally, the secreted proteins and
 CC their nucleic acids may also be used in the treatment of autoimmune
 CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
 CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
 CC wound healing. Nucleic acids of the invention may be used for chromosome
 CC identification, chromosome mapping, in gene therapy, for identifying
 CC individuals from minute biological samples, as hybridisation probes, and
 CC as molecular weight markers. The present sequence represents a human
 CC secreted protein of the invention

XX SQ Sequence 57 AA;
 Query Match 54.8%; Score 40; DB 6; Length 57;
 Best Local Similarity 57.1%; Pred. No. 1.2e+02;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 3 TNIWWWA 9
 Db 45 SHIWWWS 51
 ::|||::

RESULT 63
 ABR01346
 ID ABR01346 standard; protein; 57 AA.
 AC ABR01346;
 XX 12-MAY-2003 (first entry)
 DT Human gene 382-encoded secreted protein HLMAU42, SEQ ID NO:827.
 DE Human; secreted protein; cancer; tumour; hyperproliferative disorder;
 XX autoimmune disorder; inflammation; angiogenic diseases; AIDS;
 KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
 KW drug screening; chromosome identification; chromosome mapping;
 KW cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
 KW antianaemic; vulnery.
 XX Homo sapiens.
 OS WO200277013-A2.
 PK 03-OCT-2002.
 XX 26-MAR-2002; 2002WO-US009370.
 XX 27-MAR-2001; 2001US-0278650P.
 PR 12-SEP-2001; 2001US-00950082.
 PR 12-SEP-2001; 2001US-00950083.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Ruben SM;
 PI WPI; 2003-040578/03.
 XX N-PSDB; ABZ73680.
 DR New human secreted proteins and nucleic acids, useful for detecting or
 PT treating cancer or other hyperproliferative disorders, autoimmune
 PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia.
 XX Claim 13; Page 1507; 2474pp; English.
 PS ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted

XX ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted
 CC protein genes, and ABP00947-ABP01363 represent the proteins they encode.
 CC ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The
 CC invention also encompasses antibodies specific for the secreted proteins,
 CC the use of the secreted proteins in drug screening and recombinant
 CC vectors and host cells comprising a nucleic acid of the invention. The
 CC secreted proteins are thought to be involved in biological activities
 CC associated with cellular signalling, cellular differentiation, cell
 CC migration, prohormone activation and neurotransmitter activity. The
 CC secreted proteins, nucleic acids encoding them, antibodies or antibody
 CC fragments specific for the secreted proteins, and modulators of protein
 CC activity are useful for diagnosing or treating cancers or other
 CC hyperproliferative disorders. Additionally, the secreted proteins and
 CC their nucleic acids may also be used in the treatment of autoimmune
 CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
 CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
 CC wound healing. Nucleic acids of the invention may be used for chromosome
 CC identification, chromosome mapping, in gene therapy, for identifying
 CC individuals from minute biological samples, as hybridisation probes, and
 CC as molecular weight markers. The present sequence represents a human
 CC secreted protein of the invention
 XX
 SQ Sequence 57 AA;

Query Match 54.8%; Score 40; DB 6; Length 57;
 Best Local Similarity 57.1%; Pred. No. 1.2e+02;
 Matches 4; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 TNIIWWA 9
 : : : : :
 Db 45 SHIWWWS 51

RESULT 64
 ADC20572
 ID ADC20572 standard; protein; 57 AA.

XX AC ADC20572;

XX DT 18-DEC-2003 (first entry)

XX DE Human secreted protein - amino acid sequence #253.

XX KW gene therapy; human; secreted protein; haemopoietic disorder;
 KW haematological disorder; anaemia; haemophilia; inflammatory disorder;
 KW inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;
 KW leukaemia; wound healing; epithelial cell proliferation disorder;
 KW immune disorder; autoimmune disorder; asthmatic disorder;
 KW cardiovascular disorder; atherosclerosis; myocarditis;
 KW infectious disease; HIV; AIDS; endocrine disorder; diabetes;
 KW gastrointestinal disorder; duodenal ulcer; gastroenteritis.

XX OS Homo sapiens.

XX PN WO200292787-A2.

XX PD 21-NOV-2002.

XX PF 26-MAR-2002; 2002WO-US009257.

XX PR 27-MAR-2001; 2001US-0278650P.

XX PR 12-SEP-2001; 2001US-00950082.

XX PR 12-SEP-2001; 2001US-00950083.

XX XX (HUMA-) HUMAN GENOME SCI INC.

XX PA Rosen CA, Ruben SM;

XX PI WPI; 2003-129287/12.

XX DR New human secreted proteins and nucleic acid molecules, useful for
 XX preparing a diagnostic or pharmaceutical composition for diagnosing,

PT preventing or treating hematopoietic or hematologic disorders, e.g.
 PT anemia or hemophilia.

XX PS Claim 1; SEQ ID NO 526; 1512pp; English.

XX CC The invention comprises the amino acid and coding sequences of human
 CC secreted proteins. The DNA and protein sequences of the invention are
 CC useful for detecting, preventing, diagnosing, prognosticating, treating
 CC or ameliorating: hematopoietic or haematological disorders (e.g. anaemia
 CC and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease
 CC and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia);
 CC wound healing and disorders of epithelial cell proliferation; immune
 CC disorders (e.g. autoimmune disorders and asthmatic disorders);
 CC cardiovascular disorders (e.g. atherosclerosis and myocarditis);
 CC infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);
 CC and gastrointestinal disorders (e.g. duodenal ulcers and
 CC gastroenteritis). The present amino acid sequence represents a human
 CC secreted protein of the invention.

XX SQ Sequence 57 AA;

Query Match 54.8%; Score 43; DB 7; Length 57;

Best Local Similarity 57.1%; Pred. No. 1.2e+02;

Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 TNIIWWA 9
 : : : : :
 Db 45 SHIWWWS 51

RESULT 65

ADC20559

ID ADC20559 standard; protein; 57 AA.

XX AC ADC20559;

XX DT 18-DEC-2003 (first entry)

XX DE Human secreted protein - amino acid sequence #240.

XX KW gene therapy; human; secreted protein; haemopoietic disorder;
 KW haematological disorder; anaemia; haemophilia; inflammatory disorder;
 KW inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;
 KW leukaemia; wound healing; epithelial cell proliferation disorder;
 KW immune disorder; autoimmune disorder; asthmatic disorder;
 KW cardiovascular disorder; atherosclerosis; myocarditis;
 KW infectious disease; HIV; AIDS; endocrine disorder; diabetes;
 KW gastrointestinal disorder; duodenal ulcer; gastroenteritis.

XX OS Homo sapiens.

XX PN WO200292787-A2.

XX PD 21-NOV-2002.

XX PF 26-MAR-2002; 2002WO-US009257.

XX PR 27-MAR-2001; 2001US-0278650P.

XX PR 12-SEP-2001; 2001US-00950082.

XX PR 12-SEP-2001; 2001US-00950083.

XX XX (HUMA-) HUMAN GENOME SCI INC.

XX PA Rosen CA, Ruben SM;

XX PI WPI; 2003-129287/12.

XX DE New human secreted proteins and nucleic acid molecules, useful for
 XX preparing a diagnostic or pharmaceutical composition for diagnosing,
 XX preventing or treating hematopoietic or hematologic disorders, e.g.
 XX anemia or hemophilia.

XX PS Claim 1; SEQ ID NO 513; 1512pp; English.

XX The invention comprises the amino acid and coding sequences of human
CC secreted proteins. The DNA and protein sequences of the invention are
CC useful for detecting, preventing, diagnosing, prognosticating, treating
CC or ameliorating: haematopoietic or haematological disorders (e.g. anaemia
CC and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease
CC and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia);
CC wound healing and disorders of epithelial cell proliferation; immune
CC disorders (e.g. autoimmune disorders and asthmatic disorders);
CC cardiovascular disorders (e.g. atherosclerosis and myocarditis);
CC infectious disease (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes);
CC and gastrointestinal disorders (e.g. duodenal ulcers and
CC gastroenteritis). The present amino acid sequence represents a human
CC secreted protein of the invention.

SQ Sequence 57 AA;
Query Match 54.8%; Score 40; DB 7; Length 57;
Best Local Similarity 57.1%; Pred. No. 1.2e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 TNIIWMA 9
:|||||:
Db 45 SHIWWWS 51

RESULT 66
AAB34329
ID AAB34329 standard; protein; 58 AA.
AC AAB34329;
XX
DT 26-JAN-2001 (first entry)
XX
DE Human secreted protein sequence encoded by gene 31 SEQ ID NO:90.
XX
KW Human; secreted protein; diagnosis; neuroprotective; cytotostatic;
KW cardioactive; immunomodulatory; muscular active general; vulnary;
KW gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW and antibacterial; gene therapy; detection; cancer; chromosome marker;
KW chromosome identification; neural disorder; immune disorder;
KW muscular disorder; reproductive disorder; gastrointestinal disorder;
KW pulmonary disorder; cardiovascular disorder; renal disorder;
KW proliferative disorder; wound healing; infectious disease; preservative;
KW food additive.

XX Homo sapiens.
XX OS
XX PN WO2000056883-A1.
XX
XX PD 28-SEP-2000.
XX
XX PF 16-MAR-2000; 2000WO-US006822.
XX
XX PR 23-MAR-1999; 99US-0126054P.
XX PR 10-DEC-1999; 99US-0169916P.
XX
XX PA (HUMA-) HCVAN GENOME SCI INC.
XX
XX PI Rosen CA, Ruben SM, Komatsoulis G;
XX
XX DR WPI; 2000-587666/55.
XX DR N-PSDB; AAC59596.
XX
XX PT Human secreted proteins and gene sequences encoding them, useful for
XX PT diagnosing, preventing, and treating disorders such as cancer,
XX PT neurological disorders and immune system disorders.
XX
XX PS Claim 11; Page 380; 429pp; English.

XX The polynucleotide sequences given in AAC59566 to AAC59614 encode the
CC human secreted proteins given in AAB34299 to AAB34347. AAB34348 to
CC AAB34437 represent human secreted polypeptide sequences and proteins

CC homologous to them, which are given in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissues
CC and cells the genes are expressed in. Example of activities include:
CC neuroprotective; cytotostatic; cardioactive; immunomodulatory; muscular
CC active general; vulnary; gastrointestinal; nephrotropic; antiinfective;
CC gynaecological; and antibacterial. The polynucleotides can be used for
CC the detection of various disorders such as cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The secreted proteins can be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wound healing, and infectious diseases. The proteins can also
CC be used as a food additive or preservative to increase or decrease
CC storage capabilities. AAC59557 to AAC59565 and AAB34298 represent
CC sequences used in the exemplification of the present invention

XX SQ Sequence 58 AA;

Query Match 54.8%; Score 40; DB 3; Length 58;
Best Local Similarity 57.1%; Pred. No. 1.2e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 TNIIWMA 9
:|||||:
Db 45 SHIWWWS 51

RESULT 67
AAB51964
ID AAB51964 standard; protein; 58 AA.
AC AAB51964;
XX
XX

DT 21-FEB-2001 (first entry)

XX Human secreted protein sequence encoded by gene 36 SEQ ID NO:96.
XX
DE Human; secreted protein; immunosuppressive; antiarthritic; vulnary;
KW antirheumatic; antiproliferative; cytotostatic; cardiant; vasotrophic;
KW cerebroprotective; nontropic; neuroprotective; antibacterial; virucide;
KW fungicide; opthalmological; hyperproliferative disorder; neoplasm;
KW autoimmune disease; rheumatoid arthritis; cardiovascular disorder;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis; viral;
KW nervous system disorder; Alzheimer's disease; bacteria; infection;
KW cell proliferation; skin aging; wound healing; chemotaxis; food additive.

XX Homo sapiens.
XX OS
XX PN WO2000058334-A1.
XX
XX PD 05-OCT-2000.
XX
XX PF 22-MAR-2000; 2000WO-US007507.
XX
XX PR 26-MAR-1999; 99US-0126594P.
XX PR 17-DEC-1999; 99US-0172408P.
XX
XX PA (HUMA-) HUMAN GENOME SCI INC.

XX
XX PI Rosen CA, Ruben SM, Komatsoulis G;
XX
XX DR WPI; 2000-611701/58.
XX DR N-PSDB; AAC95497.

XX New nucleic acid molecules encoding 50 human secreted proteins for
XX PT diagnosing, preventing, treating or ameliorating medical conditions and
XX PT used as food additives or preservatives.

XX PS Claim 11; Page 382-383; 402pp; English.

XX Polynucleotide sequences AAC95462 - AAC95511 represent cDNA encoding 50
CC human secreted proteins AAB51929 - AAB51978. Sequences AAB51979 -
CC AAB52010 represent alternative polypeptides encoded by the genes, and

CC amino acid sequences with which they share homology. The genes and
 CC proteins have activities dependent on the tissues and cells in which they
 CC are expressed. Examples of their activities include immunosuppressive;
 CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
 CC vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
 CC virucide; fungicide; ophthalmological; and vulnerary. The secreted
 CC proteins, polypeptides, antagonists and agonists may be useful in
 CC treating, preventing and/or diagnosing diseases and disorders such as
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
 CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
 CC infections caused by bacteria, viruses and fungi and ocular disorders
 CC e.g. corneal infection. The polypeptides can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
 CC polypeptides can also be used as a food additive or preservative to
 CC increase or decrease storage capabilities, fat content, lipid, protein,
 CC carbohydrate, vitamins, minerals, cofactors and other nutritional
 CC components. Oligonucleotides AAC95453 - AAC95461 and peptide AAB51928 are
 CC used in the isolation and characterisation of the proteins and
 CC polynucleotides of the invention
 CC
 XX Sequence 58 AA;

Query Match 54.8%; Score 40; DB 3; Length 58;
 Best Local Similarity 57.1%; Pred. NO. 1.2e+02;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 :NIWVWA 9
 :|||||:
 Db 45 SHIWNWS 51

RESULT 68
 ABP64269
 ID ABP64269 standard; protein; 60 AA.

XX ABP64269;

AC ABP64269;

XX 04-NOV-2002 (first entry)

XX Human ORF639.

XX Cytostatic; Cardiant; Anti-allergic; Immunosuppressive; Vulnerary;
 KW Antiinflammatory; gene therapy; human; ORFX; atherogenic; platelet;
 KW human umbilical vein endothelial cell; HUVEC; atherosclerotic plaque;
 KW cancer; cardiovascular disease; allergy; autoimmune disease;
 KW wound healing; blood coagulation disorder; inflammatory disorder.

XX Homo sapiens.

XX US2002082206-A1.

XX 27-JUN-2002.

XX 30-MAY-2001; 2001US-00867550.

XX 30-MAY-2000; 2000US-0208427P.

XX (LEAC/) LEACH M D.

XX (MSHR/) MEHRABAN F.

XX (CONL/) CONLEY P B.

XX (TOPP/) TOPPER J N.

XX (LAWD/) LAW D.

XX Leach MD, Mehraban F, Conley PB, Topper JN, Law D;

XX WPI; 2002-626554/67.

XX DR N-PSDB; ABQ98832.

XX New polypeptide designated OREFX are present in human atherogenic cells

PT and are useful to prevent and treat OREFX-associated disorders including
 PT cancer, allergy, wound healing or autoimmune, cardiovascular or
 XX inflammatory disease.

PS Claim 10; SEQ ID NO 1278; 78pp; English.

XX The present invention relates to novel human OREFX polypeptides and their
 CC coding sequences (ABP63631-ABP64681 and ABQ98194-ABQ99267). The sequences
 CC were discovered in human atherogenic cells, in particular in platelets
 CC and human umbilical vein endothelial cells (HUVEC) and are expressed in
 CC many other tissues as well. Atherogenic cells are cells which have the
 CC potential to develop atherosclerotic plaques. The OREFX polypeptides and
 CC nucleic acids are useful for treating or preventing a pathological
 CC condition associated with an OREFX-associated disorder, e.g. cancer,
 CC cardiovascular disease, allergy, autoimmune disease, wound healing, blood
 CC coagulation disorders or inflammatory disorders. Note: The sequence data
 CC for this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from the USPTO web site at
 CC seqdata.uspto.gov/sequence.html?DocID=2002082206

XX Sequence 60 AA;

Query Match 54.8%; Score 40; DB 5; Length 60;
 Best Local Similarity 33.3%; Pred. NO. 1.3e+02;
 Matches 5; Conservative 2; Mismatches 2; Indels 6; Gaps 1;

QY 1 EWTNI-----WWWA 9

Db 9 EWAQVLENYVQWWS 23

RESULT 69

ABJ03768

ID ABJ03768 standard; protein; 74 AA.

XX ABJ03768;

DT 25-SEP-2002 (first entry)

XX Human ovary specific protein SEQ ID NO: 210.

XX Human; ovary; ovarian cancer; ovarian disease; gene therapy; cytostatic.

XX Homo sapiens.

XX WO2002240720-A2.

XX 23-MAY-2002.

XX 20-NOV-2001; 2001WO-US045010.

XX 20-NOV-2000; 2000US-0249997P.

XX (DIAD-) DIADEXUS INC.

XX Salceda S, Macina RA, Recipon H, Cafferkey R, Sun Y, Liu C;

XX WPI; 2002-547588/58.

XX New ovary polypeptides useful for detecting, diagnosing, monitoring,
 PT treating, staging and imaging cancers in humans having cancer and non-
 PT cancerous ovary disease.

PS Claim 11; Page 285; 296pp; English.

XX The present invention provides human proteins and coding sequences
 CC specifically found in ovary cells. These can be used in the diagnosis and
 CC treatment of ovarian diseases, including cancer. The present sequence is
 CC a protein of the invention

XX Sequence 74 AA;

Query Match 54.8%; Score 40; DB 5; Length 74;

Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 5 IWWNA 9
Db 48 VWWNA 52

RESULT 70
ABJ03783
ID ABJ03783 standard; protein; 75 AA.
AC ABJ03783;
DT 25-SEP-2002 (first entry)
DE Human ovary specific protein SEQ ID NO: 225.
KW Human; ovary; ovarian cancer; ovarian disease; gene therapy; cytostatic.
OS Homo sapiens.
XX WO200240720-A2.
XX 23-MAY-2002.
XX 20-NOV-2001; 2001WO-US045010.
XX 20-NOV-2000; 2000US-0249997P.
XX (DIAD-) DIADEXUS INC.
XX Salceda S, Macina RA, Recipon H, Cafferkey R, Sun Y, Liu C;
XX WPI; 2002-547588/58.
XX New ovary polypeptides useful for detecting, diagnosing, monitoring,
XX treating, staging and imaging cancers in humans having cancer and non-
XX cancerous ovary disease.
XX Claim 11; Page 294; 296pp; English.
XX The present invention provides human proteins and coding sequences
XX specifically found in ovary cells. These can be used in the diagnosis and
XX treatment of ovarian diseases, including cancer. The present sequence is
XX a protein of the invention
XX Sequence 75 AA;
SQ

Query Match 54.8%; Score 40; DB 5; Length 75;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 5 IWWNA 9
Db 48 VWWNA 52

RESULT 71
ABJ03781
ID ABJ03781 standard; protein; 75 AA.
AC ABJ03781;
XX 25-SEP-2002 (first entry)
DE Human ovary specific protein SEQ ID NO: 223.
KW Human; ovary; ovarian cancer; ovarian disease; gene therapy; cytostatic.
XX Homo sapiens.
XX WO200240720-A2.
PN

Query Match 54.8%; Score 40; DB 5; Length 75;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 5 IWWNA 9
Db 48 VWWNA 52

RESULT 72
ABJ03753
ID ABJ03753 standard; protein; 75 AA.
AC ABJ03753;
DT 25-SEP-2002 (first entry)
DE Human ovary specific protein SEQ ID NO: 195.
XX Human; ovary; ovarian cancer; ovarian disease; gene therapy; cytostatic.
XX Homo sapiens.
XX WO200240720-A2.
XX 23-MAY-2002.
XX 20-NOV-2001; 2001WO-US045010.
XX 20-NOV-2000; 2000US-0249997P.
XX (DIAD-) DIADEXUS INC.
XX Salceda S, Macina RA, Recipon H, Cafferkey R, Sun Y, Liu C;
XX WPI; 2002-547588/58.
XX New ovary polypeptides useful for detecting, diagnosing, monitoring,
XX treating, staging and imaging cancers in humans having cancer and non-
XX cancerous ovary disease.
XX Claim 11; Page 276-277; 296pp; English.
XX The present invention provides human proteins and coding sequences
XX specifically found in ovary cells. These can be used in the diagnosis and
XX treatment of ovarian diseases, including cancer. The present sequence is
XX a protein of the invention
XX Sequence 75 AA;
SQ

Query Match 54.8%; Score 40; DB 5; Length 75;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 5 IWWNA 9
Db 48 VWWNA 52

RESULT 73
ABJ03753
ID ABJ03753 standard; protein; 75 AA.
AC ABJ03753;
DT 25-SEP-2002 (first entry)
DE Human ovary specific protein SEQ ID NO: 195.
XX Human; ovary; ovarian cancer; ovarian disease; gene therapy; cytostatic.
XX Homo sapiens.
XX WO200240720-A2.
XX 23-MAY-2002.
XX 20-NOV-2001; 2001WO-US045010.
XX 20-NOV-2000; 2000US-0249997P.
XX (DIAD-) DIADEXUS INC.
XX Salceda S, Macina RA, Recipon H, Cafferkey R, Sun Y, Liu C;
XX WPI; 2002-547588/58.
XX New ovary polypeptides useful for detecting, diagnosing, monitoring,
XX treating, staging and imaging cancers in humans having cancer and non-
XX cancerous ovary disease.
XX Claim 11; Page 276-277; 296pp; English.
XX The present invention provides human proteins and coding sequences
XX specifically found in ovary cells. These can be used in the diagnosis and
XX treatment of ovarian diseases, including cancer. The present sequence is
XX a protein of the invention
XX Sequence 75 AA;
SQ

KW Sequence 75 AA;
 SQ Query Match 54.8%; Score 40; DB 5; Length 75;
 Best Local Similarity 80.0%; Pred. No. 1.6e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 5 IWWNA 9
 Db :|||||
 48 VWWNA 52

RESULT 73
 ABJ03740
 ID ABJ03740 standard; protein; 75 AA.
 XX AC ABJ03740;
 XX DT 25-SEP-2002 (first entry)
 XX DE Human ovary specific protein SEQ ID NO: 182.
 XX KW Human; ovary; ovarian cancer; ovarian disease; gene therapy; cytostatic.
 XX OS Homo sapiens.
 XX PN WO200240720-A2.
 XX PD 23-MAY-2002.
 XX PF 20-NOV-2000; 2001WO-US045010.
 XX PR 20-NOV-2000; 2000US-0249997P.
 XX PA (DIAD-) DIADEXUS INC.
 XX PI Salceda S, Macina RA, Recipon H, Cafferkey R, Serr Y, Liu C;
 XX DR WPI; 2002-547588/58.
 XX PT New ovary polypeptides useful for detecting, diagnosing, monitoring,
 XX PT treating, staging and imaging cancers in humans having cancer and non-
 XX PT cancerous ovary disease.
 XX PS Claim 11; Page 266-267; 296pp; English.
 XX CC The present invention provides human proteins and coding sequences
 XX CC specifically found in ovary cells. These can be used in the diagnosis and
 XX CC treatment of ovarian diseases, including cancer. The present sequence is
 XX CC a protein of the invention
 XX SQ Sequence 75 AA;
 Query Match 54.8%; Score 40; DB 5; Length 75;
 Best Local Similarity 80.0%; Pred. No. 1.6e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 5 IWWNA 9
 Db :|||||
 48 VWWNA 52

RESULT 74
 AAU60198
 ID AAU60198 standard; protein; 84 AA.
 XX AC AAU60198;
 XX DT 27-FEB-2002 (first entry)
 XX DE Propionibacterium acnes immunogenic protein #21094.
 XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;

KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.
 XX Propionibacterium acnes.
 XX WO200181581-A2.
 XX PN 01-NOV-2001.
 XX PD 20-APR-2001; 2001WO-US012865.
 XX PF 21-APR-2000; 2000US-0199047P.
 XX PR 02-JUN-2000; 2000US-0208841P.
 XX PR 07-JUL-2000; 2000US-0216747P.
 XX PA (CORI-) CORIXA CORP.
 XX PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 XX PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
 XX DR WPI; 2001-616774/71.
 XX DR N-PSDB; AAS59608.
 XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
 XX PT vaccinating against and diagnosing infections, especially useful for
 XX PT treating acne vulgaris.
 XX PS Example 1; SEQ ID NO 21393; 1069pp; English.
 XX CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 XX CC polypeptides. The proteins and their associated DNA sequences are used in
 XX CC the treatment, prevention and diagnosis of medical conditions caused by
 XX CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 XX CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 XX CC P. acnes is also involved in infections of bone, joints and the central
 XX CC nervous system, however it is particularly involved in the inflammatory
 XX CC lesions associated with acne vulgaris. A method for detecting the
 XX CC presence or absence of P. acnes in a patient comprises contacting a
 XX CC sample with a binding agent that binds to the proteins of the invention
 XX CC and determining the amount of bound protein in the sample. The
 XX CC polypeptides may be used as antigens in the production of antibodies
 XX CC specific for P. acnes proteins. These antibodies can be used to
 XX CC downregulate expression and activity of P. acnes polypeptides and
 XX CC therefore treat P. acnes infections. The antibodies may also be used as
 XX CC diagnostic agents for determining P. acnes presence, for example, by
 XX CC enzyme linked immunosorbent assay (ELISA). Note: the sequence data for
 XX CC this patent did not form part of the printed specification, but was
 XX CC obtained in electronic format directly from WIPO at
 XX CC ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 84 AA;
 Query Match 54.8%; Score 40; DB 4; Length 84;
 Best Local Similarity 66.7%; Pred. No. 1.8e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 4 NIWWNA 9
 Db :|||||
 22 SLWWNA 27

RESULT 75
 ABM56717
 ID ABM56717 standard; protein; 84 AA.
 XX AC ABM56717;
 XX DT 20-OCT-2003 (first entry)
 XX DE Propionibacterium acnes predicted ORF-encoded polypeptide #21393.
 XX KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;

immunostimulant; immune response; vaccine.

Propionibacterium acnes.

WO2003033515-A1.

24-APR-2003.

11-OCT-2002; 2002WO-US032727.

15-OCT-2001; 2001JS-00978825.

(CORI-) CORIXA CORP.

Mitcham JL, Skeiky YAM, Persing DH, Bhatia A, Maisonneuve JL; Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D; Barth B, Vallieue-Douglas J.

WPI: 2003-381789/36.

N-PSDB; ACF64537.

New Propionibacterium acnes polypeptides and polynucleotides encoding the polypeptide, useful for diagnosing, preventing or treating acne vulgaris, or for stimulating an immune response specific for a P. acnes protein.

Example 1; SEQ ID NO 21393, 1481pp; English.

The invention relates to an isolated polynucleotide (ACF64435-ACF64733) encoding a Propionibacterium acnes protein. The invention also relates to polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to immunogenic fragments of P. acnes polypeptides. The invention additionally encompasses expression vectors and host cells comprising a polynucleotide of the invention; antibodies against polypeptides of the invention; fusion proteins comprising a polypeptide of the invention; a method for stimulating an immune response specific for a P. acnes polypeptide and an isolated T cell population comprising T cells prepared via this method; a vaccine composition (comprising P. acnes polypeptides, polynucleotides, antibodies, fusion proteins, T cell populations, or antigen-presenting cells that express the polypeptide); a method and kit for detecting or determining the presence or absence of P. acnes in a patient; and a method for inhibiting the development of P. acnes in a patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion proteins, T cell populations or antigen-presenting cells that express the polypeptides are useful for diagnosing, preventing or treating acne vulgaris, or for stimulating an immune response specific for a P. acnes protein. The polynucleotides can also be used as probes or primers for nucleic acid hybridisation. The vaccine composition is useful for the stimulation of an immune response against P. acnes, or for treating acne, and the kit is useful for performing a diagnostic assay. The present sequence represents a polypeptide predicted to be encoded by an ORF (open reading frame) contained within the P. acnes polynucleotides of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences)

Sequence 84 AA;

Query Match 54.8%; Score 40; DB 6; Length 84;

Best Local Similarity 66.7%; Pred. Kc. 1.8e+02;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0

Qy 4 N1MWNNA 9

Db 22 SLMWNNA 27

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2004, 18:07:36 ; Search time 21 seconds
(without alignments)
45.805 Million cell updates/sec

Title: US-10-726-148A-15_COPY_428_437
Perfect score: 73
Sequence: 1 EWTNIWAK 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 300 summaries

Database :

PIR 78:*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	61.6	168	2 AF0632	secreted protein,
2	44	60.3	14	2 A4435	Trp EG leader pept
3	44	60.3	351	2 B46389	gene su(f) protein
4	43.5	59.6	2236	1 QZFF	rudimentary protein
5	43	58.9	95	2 E86447	protein FSD14.5 [i
6	43	58.9	688	2 T37923	hypothetical prote
7	42.5	58.2	473	2 S49939	probable membrane
8	42.5	58.2	720	2 T38647	hypothetical prote
9	42	57.5	142	2 T03976	hypothetical prote
10	42	57.5	143	2 G72000	Hth transcription
11	42	57.5	143	2 B86624	HTH transcription
12	42	57.5	144	2 B70983	hypothetical prote
13	42	57.5	319	2 B69452	hypothetical prote
14	42	57.5	352	2 A83124	dUDP-Glucose-4,6
15	42	57.5	352	2 D98163	dUDP-Glucose-4,6-d
16	42	57.5	544	2 B84264	glycine betaine tr
17	41.5	56.8	227	2 D87568	hypothetical prote
18	41.5	56.8	264	2 T36315	probable ABC-type
19	41	56.2	255	2 T52352	hypothetical prote
20	41	56.2	312	2 F90339	dUDP-Glucose 4,6-d
21	41	56.2	317	2 F90233	dUDP-Glucose 4,6-d
22	41	56.2	337	2 T18654	hypothetical prote
23	41	56.2	352	2 H87236	probable integrat
24	41	56.2	460	2 T11296	NADH2 dehydrogenas
25	41	56.2	472	2 S28286	hypothetical prote
26	41	56.2	493	2 G88553	protein C38C10.2 [
27	41	56.2	508	2 T16980	probable cytochrom
28	41	56.2	523	2 A11970	hypothetical prote
29	41	56.2	562	2 F72771	probable lysyl-TRN

30	41	56.2	692	2 E96841	hypothetical prote
31	41	56.2	764	2 C86314	hypothetical prote
32	40.5	55.5	169	2 F72532	hypothetical prote
33	40.5	55.5	589	2 H97039	phage related prot
34	40	54.8	139	2 B81941	hypothetical prote
35	40	54.8	223	2 AB2536	hypothetical prote
36	40	54.8	303	2 A64150	hypothetical prote
37	40	54.8	332	2 D69290	dUDP-glucose 4,6-d
38	40	54.8	333	2 B75098	dUDP-glucose 4,6-d
39	40	54.8	350	2 A75600	conserved hypotnet
40	40	54.8	353	2 S65297	probable membrane
41	40	54.8	359	2 E72290	branched chain ami
42	40	54.8	371	2 G69839	mandelate racemase
43	40	54.8	375	2 C87393	hypothetical prote
44	40	54.8	491	2 G89251	proline permease (
45	40	54.8	544	2 D71243	hypothetical prote
46	40	54.8	589	2 JC7520	endo-1,6-alpha-D-m
47	40	54.8	591	2 H87338	sensor histidine k
48	39.5	54.1	307	2 F95859	probable oligopept
49	39.5	54.1	2214	1 A48548	genome polyprotein
50	39	53.4	88	2 T35690	very hypothetical
51	39	53.4	133	2 A82955	hypothetical prote
52	39	53.4	157	2 B87037	probable membrane
53	39	53.4	159	2 F83733	hypothetical prote
54	39	53.4	161	2 E70530	hypothetical prote
55	39	53.4	266	2 AP3596	phosphatidylcholin
56	39	53.4	282	1 SAVL64	middle surface ant
57	39	53.4	350	2 F90334	dUDP-Glucose 4,6-d
58	39	53.4	369	2 C88030	protein F46F5.10 [
59	39	53.4	388	1 JC5461	cellulase (EC 3.2.
60	39	53.4	388	1 S43920	cellulase (EC 3.2.
61	39	53.4	412	2 C84518	hypothetical prote
62	39	53.4	414	2 S53075	probable membrane
63	39	53.4	426	1 SAVLC	large surface anti
64	39	53.4	428	1 SAVLS	large surface anti
65	39	53.4	431	1 SAVL7	large surface anti
66	39	53.4	431	1 SAVLW8	large surface anti
67	39	53.4	431	1 SAVL59	large surface anti
68	39	53.4	448	2 E96837	unknown protein T2
69	39	53.4	467	2 G90500	glycolate oxidase
70	39	53.4	512	2 G89670	glycine betaine tr
71	39	53.4	529	2 I38607	p53-binding protei
72	39	53.4	624	2 T41341	probable serine-th
73	39	53.4	642	2 G69371	acetyl-CoA synthet
74	39	53.4	748	2 S66129	disintegrin (EC 3.
75	39	53.4	863	2 JC7537	beta-N-acetylgluco
76	39	53.4	863	2 A53034	gag polyprotein -
77	39	53.4	877	2 S49197	envelope protein p
78	39	53.4	904	2 S51299	endo-beta-N-acetyl
79	39	53.4	912	2 JH0563	hypothetical prote
80	39	53.4	912	2 JH0563	metabotropic gluta
81	39	53.4	1001	2 T50914	hypothetical membr
82	38.5	52.7	282	1 WRFELS	reaction center pr
83	38.5	52.7	282	2 T50760	reaction center pr
84	38.5	52.7	572	2 B72714	probable methionyl
85	38.5	52.7	1244	2 S76102	hypothetical prote
86	38	52.1	160	2 T34665	small hydrophobic
87	38	52.1	161	2 T08983	hypothetical prote
88	38	52.1	282	2 C72282	oligopeptide ABC t
89	38	52.1	308	2 S58504	reverse transcript
90	38	52.1	328	2 D83312	hypothetical prote
91	38	52.1	336	2 S04739	site-specific DNA-
92	38	52.1	339	2 S46409	hypothetical prote
93	38	52.1	356	2 F95954	probable dUDPgluco
94	38	52.1	357	2 S22006	envelope protein g
95	38	52.1	357	2 S22004	envelope protein g
96	38	52.1	389	2 T06384	probable ethanolam
97	38	52.1	395	2 T51392	probable phytochel
98	38	52.1	433	2 T31511	hypothetical prote
99	38	52.1	460	2 A58843	NADH2 dehydrogenas
100	38	52.1	501	2 B89135	protein F25G6.7 [i
101	38	52.1	504	2 H64724	probable carnitine
102	38	52.1	504	2 C85485	probable carnitine

103	38	52.1	504	2	C90634	probable carnitine	176	37	50.7	580	2	T02596	hypothetical prote
104	38	52.1	505	2	AD0511	probable carnitine	177	37	50.7	582	2	T25837	hypothetical prote
105	38	52.1	506	2	T05156	probable glucosyl t	178	37	50.7	634	2	T49415	hypothetical prote
106	38	52.1	508	2	E90548	conserved hypotet	179	37	50.7	644	2	B87089	probable apolipop
107	38	52.1	509	2	F64210	hypothetical prote	180	37	50.7	724	2	T04340	beta-galactosidase
108	38	52.1	527	2	E87439	glycosyl transfera	181	37	50.7	772	2	JC4696	alpha,alpha-trehal
109	38	52.1	663	2	S15179	choline transport	182	37	50.7	775	2	S65769	maltooligosyl treh
110	38	52.1	677	2	H06573	high-affinity chol	183	37	50.7	792	2	T42963	hypothetical prote
111	38	52.1	677	2	D85524	high-affinity chol	184	37	50.7	818	2	T32154	hypothetical prote
112	38	52.1	747	2	AG3031	conserved hypotet	185	37	50.7	857	2	F84227	molybdenum-binding
113	38	52.1	747	2	D98254	hypothetical prote	186	37	50.7	996	1	S06635	Na+/K+-exchanging
114	38	52.1	748	2	T37097	probable secreted	187	37	50.7	1047	2	AI2002	hypothetical prote
115	38	52.1	765	2	H70763	probable gly prot	188	37	50.7	1113	2	T13348	transcription fact
116	38	52.1	856	1	VCLJVL	env polyprotein pr	189	36.5	50.0	150	2	B75528	hypothetical prote
117	38	52.1	856	1	A44963	env polyprotein pr	190	36.5	50.0	249	2	JQ2287	STCPL protein - so
118	37.5	51.4	121	2	S09959	lg heavy chain V r	191	36.5	50.0	251	2	S22202	tonoplast intrinsi
119	37.5	51.4	661	2	T22909	hypothetical prote	192	36.5	50.0	253	2	S13718	gamma tonoplast in
120	37.5	51.4	661	2	T15073	hypothetical prote	193	36.5	50.0	253	2	T51819	exodeoxyribonuclea
121	37.5	51.4	795	2	H70934	hypothetical prote	194	36.5	50.0	259	2	H81009	thiamin biosynthes
122	37.5	51.4	2164	1	GNNY89	genome polyprotein	195	36.5	50.0	283	2	T51093	hypothetical prote
123	37	50.7	45	2	C82233	hypothetical prote	196	36.5	50.0	300	2	T00274	hypothetical prote
124	37	50.7	71	2	T49130	valine-rich protei	197	36.5	50.0	725	2	T27148	H4-transporting tw
125	37	50.7	78	2	S74715	ribosomal protein	198	36	49.3	53	2	S50329	hypothetical prote
126	37	50.7	86	2	G84229	hypothetical prote	199	36	49.3	59	2	G71340	hypothetical prote
127	37	50.7	86	2	T05982	hypothetical prote	200	36	49.3	106	2	AB0256	probable phage pro
128	37	50.7	104	2	D72542	hypothetical prote	201	36	49.3	124	2	AG2441	hypothetical prote
129	37	50.7	106	2	T36397	hypothetical prote	202	36	49.3	132	2	E70074	hypothetical prote
130	37	50.7	116	2	G97237	probable integral	203	36	49.3	171	2	T11285	NADH2 dehydrogenas
131	37	50.7	126	2	T47859	probable membrane	204	36	49.3	183	2	AI0509	hypothetical prote
132	37	50.7	154	2	JC6036	hypothetical prote	205	36	49.3	210	2	B40203	4-alpha-glucanotra
133	37	50.7	160	2	C69300	hypothetical prote	206	36	49.3	222	2	H90978	hypothetical prote
134	37	50.7	190	2	D84942	translation initia	207	36	49.3	242	2	AG1868	hypothetical prote
135	37	50.7	194	2	D75078	hypothetical prote	208	36	49.3	256	2	T10821	photosynthetic rea
136	37	50.7	205	2	S74960	hypothetical prote	209	36	49.3	284	2	AD3353	probable membrane
137	37	50.7	209	2	T44553	conserved hypotet	210	36	49.3	275	2	T32813	hypothetical prote
138	37	50.7	212	2	S77369	hypothetical prote	211	36	49.3	279	2	A96913	ABC-type sulfate t
139	37	50.7	220	2	AH2270	hypothetical prote	212	36	49.3	293	2	E83513	probable outer mem
140	37	50.7	229	2	F87385	hypothetical prote	213	36	49.3	302	2	B75162	ribose abc transpo
141	37	50.7	217	2	T05249	hypothetical prote	214	36	49.3	302	2	B71179	hypothetical prote
142	37	50.7	249	2	T19088	hypothetical prote	215	36	49.3	325	2	C75280	phenylacetic acid
143	37	50.7	277	2	T29611	hypothetical prote	216	36	49.3	330	2	H72588	probable drpp-gluc
144	37	50.7	283	2	T22048	hypothetical prote	217	36	49.3	332	2	G96036	probable phenylace
145	37	50.7	298	2	B90435	hypothetical prote	218	36	49.3	344	2	D95158	conserved domain p
146	37	50.7	298	2	A89009	protein T27C4.1 [i	219	36	49.3	344	2	D98024	hypothetical prote
147	37	50.7	306	2	B75142	oligopeptide trans	220	36	49.3	352	2	S45558	cytochrome c-type
148	37	50.7	312	2	D83239	probable lauroyl a	221	36	49.3	355	2	A23604	cytochrome d ubiq
149	37	50.7	317	2	T35010	probable integral	222	36	49.3	357	2	S21990	envelope protein g
150	37	50.7	318	2	E82493	conserved hypotet	223	36	49.3	361	1	F69254	probable hexosyltr
151	37	50.7	323	2	G90487	maltose ABC transp	224	36	49.3	363	2	F69325	first mannosyl tra
152	37	50.7	334	2	AC0822	probable DNA-bind	225	36	49.3	364	2	C75597	thymidine diphosph
153	37	50.7	334	2	A96700	protein F12A21.19	226	36	49.3	393	2	E82283	conserved hypotet
154	37	50.7	336	2	G71151	probable drpp-gluc	227	36	49.3	394	2	JC7639	chalcone synthase-
155	37	50.7	345	2	AR0350	probable DNA-bind	228	36	49.3	396	2	C82769	phage-related cont
156	37	50.7	356	2	B75438	GGDEF family prote	229	36	49.3	399	2	C83847	cytochrome c bioge
157	37	50.7	362	2	T21514	hypothetical prote	230	36	49.3	409	2	T50311	oxalylp2 protein [i
158	37	50.7	377	2	B83454	probable alkane hy	231	36	49.3	409	2	T43703	respiratory protei
159	37	50.7	377	2	T36322	probable membrane	232	36	49.3	426	2	A42649	cellulase [EC 3.2
160	37	50.7	388	2	T09885	hypothetical prote	233	36	49.3	432	2	A85040	hypothetical prote
161	37	50.7	445	2	S77010	nickel resistance	234	36	49.3	435	2	T47737	hypothetical prote
162	37	50.7	451	1	S49016	protein-tyrosine k	235	36	49.3	460	2	T09956	NADH2 dehydrogenas
163	37	50.7	452	2	T35729	hypothetical prote	236	36	49.3	466	2	S09837	hypothetical prote
164	37	50.7	454	2	B41621	env polyprotein D	237	36	49.3	472	2	T41684	probable sterol o-
165	37	50.7	460	2	T11309	NADH2 dehydrogenas	238	36	49.3	477	1	B64813	ybH1 protein - Esc
166	37	50.7	463	2	AB2543	nickel resistance	239	36	49.3	477	2	G69760	beta-glucosidase h
167	37	50.7	486	2	AF1174	-lysine-specific pe	240	36	49.3	477	2	F90728	probable membrane
168	37	50.7	486	2	AG1531	-lysine-specific pe	241	36	49.3	477	2	G85579	ammonium transport
169	37	50.7	490	2	F82546	fimbrial assembly	242	36	49.3	510	2	C87415	phenol 2-monooxyge
170	37	50.7	491	2	JE0276	voltage-gated pota	243	36	49.3	511	1	S47290	phenol 2-monooxyge
171	37	50.7	491	2	C84362	Na+/H+ antiporter	244	36	49.3	513	2	T01506	probable hexose tr
172	37	50.7	500	1	Q08868	HHLF6 protein - hu	245	36	49.3	513	2	T41011	hypothetical prote
173	37	50.7	515	2	G70941	hypothetical prote	246	36	49.3	514	2	T40485	transmembrane tran
174	37	50.7	516	2	H82973	choline transporte	247	36	49.3	518	1	G69804	multidrug-efflux t
175	37	50.7	531	2	T32276	hypothetical prote	248	36	49.3	521	2	A83451	aerotaxis receptor

249 36 49.3 523 2 S25015 monosaccharide tra
 250 36 49.3 526 2 T01853 probable hexose tr
 251 36 49.3 547 2 E70720 probable lrvG prot
 252 36 49.3 548 2 T44762 probable rfp-regul
 253 36 49.3 570 2 S04547 proline transport
 254 36 49.3 570 2 S42708 proline transport
 255 36 49.3 580 2 S76846 hypothetical prote
 256 36 49.3 665 2 H96582 F15t1.27 imported
 257 36 49.3 673 2 AF0143 high-affinity chol
 258 36 49.3 746 1 HYH0MA meprin A (EC 3.4.2
 259 36 49.3 748 2 S24134 endopeptidase 2 (E
 260 36 49.3 757 2 JC7726 (1->4)-alpha-D-glu
 261 36 49.3 758 2 T48815 mixed-linked glucu
 262 36 49.3 760 2 A40195 meprin A (EC 3.4.2
 263 36 49.3 766 2 H81141 probable two-compo
 264 36 49.3 768 2 JC8564 cellobiose oxidase
 265 36 49.3 797 2 A36811 hypothetical prote
 266 36 49.3 836 2 JE0248 ATP-binding casset
 267 36 49.3 860 2 C72339 hypothetical prote
 268 36 49.3 868 1 VCLJH4 env polypeptid -
 269 36 49.3 943 2 T35497 hypothetical prote
 270 36 49.3 964 2 T41547 hypothetical prote
 271 36 49.3 984 2 E70406 DMSO reductase cha
 272 36 49.3 996 2 A71080 hypothetical prote
 273 36 49.3 1007 2 H72734 hypothetical prote
 274 36 49.3 1010 2 F75134 hypothetical prote
 275 36 49.3 1020 2 D83679 hypothetical prote
 276 36 49.3 1307 2 T17453 ERG-associated pro
 277 36 49.3 1374 2 T30809 plasmidogen relate
 278 36 49.3 1515 2 A40203 4-alpha-glucanotra
 279 36 49.3 1555 2 S38758 amylo-alpha-1,6-gl
 280 36 49.3 1779 2 T23130 hypothetical prote
 281 36 49.3 2078 2 T35400 hypothetical prote
 282 36 49.3 2895 2 T08437 hyperplastic discs
 283 35.5 48.6 349 2 S53864 hypothetical prote
 284 35.5 48.6 718 1 S33168 gene pointed prote
 285 35.5 48.6 977 2 T41289 hypothetical prote
 286 35.5 48.6 1011 2 T40851 hypothetical prote
 287 35 47.9 84 2 S76840 hypothetical prote
 288 35 47.9 151 2 S30457 env protein - huma
 289 35 47.9 151 2 S30456 env protein - huma
 290 35 47.9 151 2 S30455 env protein - huma
 291 35 47.9 151 2 S30454 env protein - huma
 292 35 47.9 177 2 C69137 hypothetical prote
 293 35 47.9 178 2 H83349 hypothetical prote
 294 35 47.9 188 2 A75382 probable alanine-r
 295 35 47.9 189 2 T36207 hypothetical prote
 296 35 47.9 192 2 AD2294 hypothetical prote
 297 35 47.9 196 2 B71232 transcription regu
 298 35 47.9 199 2 C75449 NAD(P)H oxidoreduc
 299 35 47.9 202 2 E84135 acetyltransferase
 300 35 47.9 219 2 JN0822

ALIGNMENTS

RESULT 1

AF0632
 A:Title: secreted protein, suppressor for copper-sensitivity D [imported] - Salmonella enterica s
 C:Species: Salmonella enterica subsp. enterica serovar Typhi
 A:Note: This species has also been called Salmonella typhi
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C:Accession: AF0632
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, M.R.; Pickard, D.; Main, J.; Churcher,
 th, T.; Connor, P.; Cronin, P.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
 S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
 A:Reference number: AB0502; MUID:21534947; PMID:11677608
 A:Accession: AF0632
 A>Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-168 <PAR>
 A:Cross-references: GS:AL513382; PIDN:CAD08241.1; PID:G16502288; GSPDB:GN00176
 C:Genetics:
 A:Gene: scsd
 Query Match 61.6%; Score 45; DB 2; Length 168;
 Best Local Similarity 46.7%; Pred. No. 13;
 Matches 7; Conservative 1; Mismatches 1; Indels 6; Gaps 1;
 QY 2 WTNIW-----WWAK 10
 ||:|
 DB 152 WTSYGMKRLWWAK 166
 RESULT 2
 A44515
 Trp EG leader peptide - Rhizobium meliloti
 C:Species: Rhizobium meliloti
 C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 12-Dec-1997
 C:Accession: A44515
 R:Bae, Y.M.; Holmgren, E.; Crawford, I.P.
 J. Bacteriol. 171, 3471-3478, 1989
 A:Title: Rhizobium meliloti anthranilate synthase gene: cloning, sequence, and expressi
 A:Reference number: A44515; MUID:89255120; PMID:2656657
 A:Accession: A44515
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-14 <BAE>
 A:Cross-references: EMBL:M22983
 C:Superfamily: unassigned leader peptides
 Query Match 60.3%; Score 44; DB 2; Length 14;
 Best Local Similarity 71.4%; Pred. No. 1.5;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 4 NIWWAK 10
 :|||:
 DB 3 SIWWAR 14
 RESULT 3
 B46389
 gene su(f) protein, 42K splice form - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
 C:Accession: B46389
 R:Mitchelson, A.; Simonelig, M.; Williams, C.; O'Hare, K.
 Genes Dev. 7, 241-249, 1993
 A:Title: Homology with Saccharomyces cerevisiae RNA14 suggests that phenotypic suppress
 A:Reference number: A46389; MUID:93170663; PMID:8436295
 A:Accession: B46389
 A>Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-351 <MIT>
 A:Cross-references: GB:S55230; GB:S55231; GB:S55232; GB:S55233; MUID:92511641
 A:Note: sequence extracted from NCBI backbone {NCBIN:125337, NCBIN:125340, NCBI:125358
 C:Genetics:
 A:Gene: FlyBase:su(f)
 A:Cross-references: FlyBase:FBgn0003559
 Query Match 60.3%; Score 44; DB 2; Length 351;
 Best Local Similarity 58.3%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 3; Indels 2; Gaps 1;
 QY 1 EWTN--IWWAK 10
 ||
 DB 334 EWWAFANWWAK 345
 RESULT 4
 Q0FF
 rudimentary protein - fruit fly (Drosophila melanogaster)

N:Alternate names: CAD protein
N:Contains: aspartate carbamoyltransferase (EC 2.1.3.2); carbamoyl-phosphate synthase (C)
C:Species: Drosophila melanogaster
C:Date: 31-Mar-1989 #sequence_revision 23-Feb-1996 #text_change 18-Jun-1999
C:Accession: A29106; S53752; S35447; S35446
R:Freund, J.N.; Jarry, B.P.
J. Mol. Biol. 193, 1-13, 1987
A:Title: The rudimentary gene of Drosophila melanogaster encodes four enzymic functions.
A:Reference number: A29106; MUID:87226179; PMID:2884325
A:Accession: A29106
A:Molecule type: mRNA
A:Residues: 1-2074; 'WRLERTDRALAGPPADPVQCEPAVMRRD', 2105-2130, 'ECAAHHGSCALHDSHS', 2146-2233
A:Cross-references: GB:X04813; NID:G8508; EIDN:CAA28502.1; PID:G8509
A:Note: this sequence has been revised in reference S53752
R:Davidson, J.N.; Kern, C.B.
J. Mol. Biol. 243, 364-366, 1994
A:Title: Revision in sequence of CAD aspartate transcarbamylase domain of Drosophila.
A:Reference number: S53752; MUID:95018278; PMID:7932764
A:Accession: S53752
A:Molecule type: mRNA
A:Residues: 2068-2148 <DAV>
A:Cross-references: GB:S74010; NID:G693827; PIDN:AB32204.1; PID:G693828
A:Note: this is a revision to the sequence from reference A29106
R:Zerges, B.; Udvady, A.; Schedl, P.
Nucleic Acids Res. 20, 4639-4647, 1992
A:Title: Molecular characterization of the 5' end of the rudimentary gene in Drosophila
A:Reference number: S35446; MUID:93027163; PMID:1329025
A:Accession: S35446
A:Molecule type: DNA
A:Residues: 1-109 <ZER2>
A:Cross-references: EM3L:M37783
C:Genetics:
A:Gene: FlyBase:r
A:Cross-references: FlyBase:FBgn0003189
C:Superfamily: rudimentary enzyme; aspartate/ornithine carbamoyltransferase homology; Ba
arbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain homology; carbamoyl-phos
C:Keywords: hydrolase; ligase; multifunctional enzyme; pyrimidine nucleotide biosynthesi
F:8-1450/Domain: carbamoyl-phosphate synthase (ammonia) homology <CPA>
F:8-380/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) small chain homolog
F:201-376/Domain: trpG homology <TRG>
F:408-865/Domain: biotin carboxylase homology <BC1>
F:410-1457/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain hom
F:953-1396/Domain: biotin carboxylase homology <BC2>
F:1475-1818/Domain: Bacillus dihydroxyotase homology <DHC>
F:1923-2219/Domain: aspartate/ornithine carbamoyltransferase homology <ACT>
F:274/Active site: Cys #status predicted

Query Match 59.6%; Score 43.5; DB 1; Length 2236;
Best Local Similarity 41.2%; Pred. No. 2.7e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 7; Gaps 1;
Qy 1 EWTN-----IWWAK 10
|||
Db 88 EWTEGREGPPPPWWAR 104

RESULT 5
S86447
protein F5D14.5 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: S86447
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
ansen, N.F.; Hughes, B.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
Nature 408, 85-6-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzial
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallor
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: S86447
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-95 <STO>
A:Cross-references: GB:AE005172; NID:G8920603; PIDN:AAF81325.1; GSPDB:GN00141
C:Genetics:
A:Gene: F5D14.5
A:Map position: 1

Query Match 58.9%; Score 43; DB 2; Length 95;
Best Local Similarity 62.5%; Pred. No. 14;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 EWTNIIWWM 8
|||
Db 75 EWSWWWWWW 82

RESULT 6
T37923
hypothetical protein SPAC18G6.10 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T37923
R:Conor, R.; Churcher, C.M.; Barrrell, B.G.; Rajandream, M.A.; Walsb, S.V.; Wood, V.
submitted to the EMBL Data Library, December 1995
A:Reference number: 221754
A:Accession: T37923
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-688 <CON>
A:Cross-references: EMBL:Z68198; PIDN:CAA92388.1; GSPDB:GN00066; SPDB:SPAC18G6.10
A:Experimental source: strain 972h-; cosmid cl866
C:Genetics:
A:Map position: 1

Query Match 58.9%; Score 43; DB 2; Length 688;
Best Local Similarity 62.5%; Pred. No. 98;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 EWTNIIWWM 8
|||
Db 658 EWTRVWEM 665

RESULT 7
S49939
probable membrane protein YII039w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YI9905.09
C:Species: Saccharomyces cerevisiae
C:Date: 13-Jan-1995 #sequence_revision 20-Feb-1995 #text_change 19-Apr-2002
C:Accession: S49939
R:Odell, C.; Bowman, S.
submitted to the EMBL Data Library, December 1994
A:Reference number: S49931
A:Accession: S49939
A:Molecule type: DNA
A:Residues: 1-473 <ODE>
A:Cross-references: GB:Z47047; EMBL:Z46861; NID:G603997; PID:G763307; GSPDB:GN00009; N
C:Genetics:
A:Gene: MIPS:YII039w
A:Cross-references: SGD:S0001301
A:Map position: 9L
C:Keywords: transmembrane protein
F:457-473/Domain: transmembrane #status predicted <TMM>

Query Match 58.2%; Score 42.5; DB 2; Length 473;
 Best Local Similarity 36.8%; Pred. No. 79;
 Matches 7; Conservative 3; Mismatches 0; Indels 9; Gaps 1;

QY 1 EWT-----NINWMAK 10
 |||
 439 EWTFSLOPFAIQHWNFAK 457
 |||

Db

RESULT 8
 T38647
 hypothetical protein SPAC32A11.01 - fission yeast (*Schizosaccharomyces pombe*)
 C:Species: *Schizosaccharomyces pombe*
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
 C:Accession: T38647
 R:Skellton, J.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
 submitted to the EMBL Data Library, February 1996
 A:Reference number: Z21803
 A:Accession: T38647
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-720 <SK>
 A:Cross-references: EMBL:Z69796; PIDN:CAA91698.1; GSPDB:GN00066; SPDB:SPAC32A11.01
 A:Experimental source: strain 972h; cosmid c32A11
 C:Genetics:
 A:Gene: SPDB:SPAC32A11.01
 A:Map position: 1
 A:Introns: 45/1; 210/2
 C:Superfamily: *Schizosaccharomyces* hypothetical protein SPAC32A11.01

Query Match 58.2%; Score 42.5; DB 2; Length 720;
 Best Local Similarity 77.8%; Pred. No. 1.2e+02;
 Matches 7; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 3 TINIWW-WAK 10
 |||||
 398 TINIWWNAE 406
 |||||

Db

RESULT 9
 T03976
 hypothetical protein - maize insertion element Bsl
 C:Species: *Zea mays* (maize)
 C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
 C:Accession: T03976
 R:Jin, Y.K.; Bennett, J.L.
 Proc. Natl. Acad. Sci. U.S.A. 86, 6235-6239, 1989
 A:Title: Structure and coding properties of Bsl, a maize retrovirus-like transposon.
 A:Reference number: Z15172; MUID:89345638; PMID:2474829
 A:Accession: T03976
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-142 <JIN>
 A:Cross-references: EMBL:M25397; NID:g168648; PIDN:AAA66270.1; PID:g806302
 C:Genetics:
 A:Mobile element: insertion element Bsl

Query Match 57.5%; Score 42; DB 2; Length 142;
 Best Local Similarity 71.4%; Pred. No. 28;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 WTNIWW 8
 :|||
 68 FTNPWW 74
 :|||

Db

RESULT 10
 G72000
 Hth transcription regulator - *Chlamydomonas reinhardtii* (strains CW029 and AR39)
 C:Species: *Chlamydomonas reinhardtii*, *Chlamydomonas reinhardtii*
 C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
 C:Accession: G72000; B81538

R.Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.
 Nature Genet. 21, 385-389, 1999
 A:Title: Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*.
 A:Reference number: A72000; MUID:99206606; PMID:10192388
 A:Accession: G72000
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-143 <ARN>
 A:Cross-references: GB:AE001687; GB:AE001363; NID:g4377398; PIDN:AA19206.1; PID:g4377
 A:Experimental source: strain CW029
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Gwinn, W.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39.
 A:Reference number: A81500; MUID:20150255; PMID:10684935
 A:Accession: B81538
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-143 <REA>
 A:Cross-references: GB:AE002238; GB:AE002161; NID:g7189593; PIDN:AAF38580.1; PID:g7189
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: yfgA; CP0781

Query Match 57.5%; Score 42; DB 2; Length 143;
 Best Local Similarity 66.7%; Pred. No. 28;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 WTNIWW 7
 |||
 117 WSNLWW 122
 |||

Db

RESULT 11
 B86624
 HTH transcription regulator [imported] - *Chlamydomonas reinhardtii* (strain J138)
 C:Species: *Chlamydomonas reinhardtii*, *Chlamydomonas reinhardtii*
 C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
 C:Accession: B86624
 R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;
 Nucleic Acids Res. 28, 2311-2314, 2000
 A:Title: Comparison of whole genome sequences of *Chlamydia pneumoniae* J138.
 A:Reference number: A86491; MUID:20330349; PMID:10871362
 A:Accession: B86624
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-143 <STO>
 A:Cross-references: GB:BA000008; NID:g8979442; PIDN:BAA99276.1; GSPDB:GN00142
 A:Experimental source: strain J138
 C:Genetics:
 A:Gene: yfgA

Query Match 57.5%; Score 42; DB 2; Length 143;
 Best Local Similarity 66.7%; Pred. No. 28;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 WTNIWW 7
 |||
 117 WSNLWW 122
 |||

Db

RESULT 12
 B70983
 hypothetical protein Rv0121c - *Mycobacterium tuberculosis* (strain H37RV)
 C:Species: *Mycobacterium tuberculosis*
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: B70983
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
 Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome

A;Reference number: A7C500; MUID:98295987; PMID:9634230
A;Accession: B70983
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-144 <COL>
A;Cross-references: GB:Z9607L; GB:AL123456; NID:g3242254; PIDN:CAB09449.1; PID:g2181963
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: Rv0121c
C;Superfamily: Mycobacterium tuberculosis hypothetical protein Rv0121c

Query Match 57.5%; Score 42; DB 2; Length 144;
Best Local Similarity 57.1%; Pred. No. 28;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNIWW 7
: : : : :
86 DWTQLWN 92

Db

RESULT 13
B69452
hypothetical protein AF1619 - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C;Accession: B69452
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glocke, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A;Reference number: A93250; MUID:98049343; PMID:9389475
A;Accession: B69452
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-319 <KLE>
A;Cross-references: GB:AE000991; GB:AE000782; NID:g2689314; PIDN:AA889643.1; PID:g264895

Query Match 57.5%; Score 42; DB 2; Length 319;
Best Local Similarity 38.5%; Pred. No. 63;
Matches 5; Conservative 4; Mismatches 0; Indels 4; Gaps 1;

QY 1 EWTNIWW----WVA 9
: : : : :
8 DWSSLWXEDWVA 20

Db

RESULT 14
AE3124
dUDP-D-glucose-4,6-dehydratase rffb [imported] - Agrobacterium tumefaciens (strain C58,
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AE3124
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuryavin, T.; Levy, R.; Ji, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AE3124
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-352 <KUR>
A;Cross-references: GB:AE000689; PIDN:AA145411.1; PID:g1774311C; GSPDB:GN00187
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: rffb
A;Map position: linear chromosome
C;Superfamily: Escherichia coli UDPglucose 4-epimerase; UDPglucose 4-epimerase homology

Query Match 57.5%; Score 42; DB 2; Length 352;
Best Local Similarity 60.0%; Pred. No. 69;
Matches 6; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

QY 1 EW--TNIWW 8
: : : : :
324 EWYLDNAWW 333

Db

RESULT 15
D98163
dtdp-glucose-4,6-dehydratase (AE314183) [imported] - Agrobacterium tumefaciens (strain
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C;Accession: D98163
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quicillo, B.; Goldma
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, I
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: D98163
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-352 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK88830.1; PID:gl5158590; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_530
A;Map position: linear chromosome
C;Superfamily: Escherichia coli UDPglucose 4-epimerase; UDPglucose 4-epimerase homolo

Query Match 57.5%; Score 42; DB 2; Length 352;
Best Local Similarity 60.0%; Pred. No. 69;
Matches 6; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

QY 1 EW--TNIWW 8
: : : : :
324 EWYLDNAWW 333

Db

RESULT 16
B84264
glycine betaine transporter [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: B84264
R;NG, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky,
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jal
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omar, A.D.; Ebhardt, H.; Lowe, T.M.;
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: B84264
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-544 <STO>
A;Cross-references: GB:AE004437; NID:gl0580626; PIDN:AAGL19478.1; GSPDB:GN00138
C;Genetics:
A;Gene: opud

Query Match 57.5%; Score 42; DB 2; Length 544;
Best Local Similarity 60.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

QY 2 WTNIW--WVA 9
: : : : :
325 WTGFVWAWVA 334

Db

RESULT 17
D87568
hypothetical protein CC2574 [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus.
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: D87568
R:NIEMAN, M.C.; FELDBLYUM, T.V.; PAULSEN, I.T.; NELSON, K.E.; RISEN, J.; HEIDELBERG, J.
B.; LAUB, M.T.; DEBOY, R.T.; DODSON, R.J.; DURKIN, A.S.; GWINN, M.L.; HAFT, D.H.; KOLON
N, J.; ERMOLAeva, M.; WHITE, O.; SALZBERG, S.L.; SHAPIRO, L.; VENTER, J.C.; FRASER, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: D87568
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-227 <STO>
A:Cross-references: GB:AE005673; NID:gl3424142; PIDN:AAK24544.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2574

Query Match 56.8%; Score 41.5; DB 2; Length 227;
Best Local Similarity 45.5%; Pred. No. 52;
Matches 5; Conservative 3; Mismatches 0; Indels 3; Gaps 1;

QY 2 WTNINW---WA 9
||:|:| |
Db 208 WSSVWMAAWA 218

RESULT 18
T36315
probable ABC-type transport system integral membrane protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36315
R:Saunders, D.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, March 1999
A:Reference number: Z21604
A:Accession: T36315
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-264 <SAU>
A:Cross-references: EMBL:AL035654; PIDN:CAB38594.1; GSPDB:GN00070; SCOEDB:SCE8.16c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCE8.16c

Query Match 56.8%; Score 41.5; DB 2; Length 264;
Best Local Similarity 46.2%; Pred. No. 61;
Matches 6; Conservative 0; Mismatches 2; Indels 5; Gaps 1;

QY 1 EWTNINW-----WW 8
||| | |
Db 25 EWTKFSVRATWW 37

RESULT 19
T52352
hypothetical protein B11E6.10 [imported] - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C:Accession: T52352
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, September 2000
A:Reference number: Z26053
A:Accession: T52352
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-255 <SCH>
A:Cross-references: EMBL:AL442043; GSPDB:GN00116; NCSP:B11E6.10
A:Experimental source: BAC clone B11E6; strain OR74A
C:Genetics:
A:Gene: NCSP:B11E6.10
A:Map position: 6

Query Match 56.2%; Score 41; DB 2; Length 255;

C;Accession: T18654
 R;White, S.
 submitted to the EMBL Data Library, May 1996
 A;Reference number: Z19002
 A;Accession: T18654
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-337 <WIL>
 A;Cross-references: EMBL:Z73102; PIDN:CAA97409.1; GSPDB:GN00022; CRSP:B0035.2
 C;Genetics:
 A;Gene: CRSP:B0035.2
 A;Map position: 4
 A;Introns: 40/1; 111/2; 177/3; 210/2; 276/3

Query Match 56.2%; Score 41; DB 2; Length 337;
 Best Local Similarity 50.0%; Pred. No. 90;
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNIWW 8
 Db 238 KWTALWY 245
 :||:|:|:

RESULT 23
 H87236
 Probable integral membrane protein [imported] - Mycobacterium leprae
 C;Species: Mycobacterium leprae
 C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C;Accession: H87236
 R;Cole, S.T.; Eiglmier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Holt,
 R.; Davies, R.V.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
 ean, M.A.; Rutherford, K.M.
 Nature 409, 1007-1011, 2001
 A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
 A;Title: Massive gene decay in the leprosy bacillus.
 A;Reference number: A86909; MUID:21128732; PMID:11234002
 A;Accession: H87236
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-352 <STO>
 A;Cross-references: GB:AL450380; NID:g13093844; PIDN:CAC32150.1; GSPDB:GN00147
 C;Genetics:
 A;Gene: ML2618

Query Match 56.2%; Score 41; DB 2; Length 352;
 Best Local Similarity 45.5%; Pred. No. 94;
 Matches 5; Conservative 2; Mismatches 0; Indels 4; Gaps 1;

QY 2 WTNI---WW 8
 Db 44 WTSVYEW 54
 :||:|:|:

RESULT 24
 T11296
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - Raja radiata mitochondrion
 C;Species: mitochondrion Raja radiata
 C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 03-Jun-2002
 C;Accession: T11296
 R;Rasmussen, A.S.; Arnason, U.
 Proc. Natl. Acad. Sci. U.S.A. 96, 2177-2182, 1999
 A;Title: Molecular studies suggest that cartilaginous fishes have an apical position in
 A;Reference number: Z17259; MUID:99162577; PMID:10051614
 A;Accession: T11296
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-460 <RAS>
 A;Cross-references: EMBL:AF106038; NID:g4406269; PIDN:G4406279; PIDN:AA19338.1
 C;Genetics:
 A;Gene: NADH4
 A;Genome: mitochondrion
 A;Genetic code: SGCI

C;Superfamily: NADH dehydrogenase [ubiquinone] chain 4
 C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation

Query Match 56.2%; Score 41; DB 2; Length 460;
 Best Local Similarity 62.5%; Pred. No. 1.2e+02;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 WTNIIWW 9
 Db 191 WANKFWMA 198
 :||:|:|:

RESULT 25
 S28286
 Hypothetical protein C38C10.2 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 16-Feb-1997
 C;Accession: S28286
 R;Thomas, K.
 submitted to the EMBL Data Library, December 1992
 A;Reference number: S28286
 A;Accession: S28286
 A;Molecule type: DNA
 A;Residues: 1-472 <THO>
 A;Cross-references: EMBL:Z19153
 C;Genetics:
 A;Introns: 50/3; 287/3; 351/3; 412/3
 A;Keywords: transmembrane protein

Query Match 56.2%; Score 41; DB 2; Length 472;
 Best Local Similarity 57.1%; Pred. No. 1.3e+02;
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WTNIIWW 8
 Db 197 WTAVMWY 203
 :||:|:|:

RESULT 26
 G88553
 protein C38C10.2 [imported] - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
 C;Accession: G88553
 R;anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998
 A;Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
 A;Reference number: A75000; MUID:99069613; PMID:9851916
 A;Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C_e
 A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; e
 A;Accession: G88553
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-493 <STO>
 A;Cross-references: GB:chr_III; PIDN:CAA79549.1; PID:g3874873; GSPDB:GN00021; CRSP:C38
 C;Genetics:
 A;Gene: C38C10.2
 A;Map position: 3

Query Match 56.2%; Score 42; DB 2; Length 493;
 Best Local Similarity 57.1%; Pred. No. 1.3e+02;
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WTNIIWW 8
 Db 218 WTAVMWY 224
 :||:|:|:

RESULT 27
 T16980
 Probable cytochrome P-450 - curled-leaved tobacco
 N;Contains: probable oxidoreductase (EC 1.-.-.-)
 C;Species: Nicotiana glauca (curled-leaved tobacco)

C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 16-Feb-2001
 C;Accession: T:6980
 R;Lakosa, P.C.; Smigocki, A.C.
 submitted to the EMBL Data Library, March 1999
 A;Description: A wound-inducible cytochrome P-450 in *Nicotiana glauca* plumbaginifolia.
 A;Reference number: 218630
 A;Accession: T:6980
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-508 <LAR>
 A;Cross-references: EMBL:U35226; NID:94376202; PID:94376203
 A;Experimental source: shoots (stems and leaves); preflowering rosettes
 C;Superfamily: human cytochrome P450 CYP4B1; cytochrome P450 homolog
 C;Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase;
 F;317-478/Domain: cytochrome P450 homolog <P45>
 F;456/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 56.2%; Score 41; DB 2; Length 508;
 Best Local Similarity 83.3%; Pred. No. 1.4e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 IWWAK 10
 Db 26 IWWPK 31
 |||||

RESULT 28
 All1970
 Hypothetical protein all1316 [imported] - Nostoc sp. (strain PCC 7120)
 C;Species: Nostoc sp. PCC 7120
 A;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C;Accession: All1970
 R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
 Nakazaki, N.; Shampo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Ana*
 A;Reference number: AB1807; MUID:21595285; PMID:11759840
 A;Accession: All1970
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-523 <KUR>
 A;Cross-references: GB:BA000019; PIDN:BA073273.1; PID:gi17130663; GSPDB:GN00179
 A;Experimental source: strain PCC 7120
 C;Genetics:
 A;Gene: all1316

Query Match 56.2%; Score 41; DB 2; Length 523;
 Best Local Similarity 80.0%; Pred. No. 1.4e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 4 NIWW 8
 Db 198 NLWW 202
 |||||

RESULT 29
 F72771
 Probable lysyl-tRNA Synthetase AP00161 - *Aeropyrum pernix* (strain K1)
 C;Species: *Aeropyrum pernix*
 C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
 C;Accession: F72771
 R;Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
 DNA Res. 6, 83-101, 1999
 A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, *Aeropyr*
 A;Reference number: A72450; MUID:99310339; PMID:10382966
 A;Accession: F72771
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-562 <KAW>
 A;Cross-references: DDBJ:AP000058; NID:95103388; PIDN:BA079072.1; PID:95103351
 A;Experimental source: strain K1

C;Genetics:
 A;Gene: AP00161
 C;Superfamily: Lyme disease spirochete lysine-tRNA ligase

Query Match 56.2%; Score 41; DB 2; Length 562;
 Best Local Similarity 57.1%; Pred. No. 1.5e+02;
 Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EWTNIWW 7
 Db 246 EWAGVWW 252
 |||||

RESULT 30
 E36841
 Hypothetical protein F23A5.23 [imported] - *Arabidopsis thaliana*
 C;Species: *Arabidopsis thaliana* (mouse-ear cress)
 C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C;Accession: E36841
 R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alons
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, I
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzial
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A;Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
 A;Reference number: AB6141; MUID:21016719; PMID:11130712
 A;Accession: E36841
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-692 <STO>
 A;Cross-references: GB:AE005173; NID:96503299; PIDN:AAF14675.1; GSPDB:GN00141
 C;Genetics:
 A;Gene: F23A5.23
 A;Map position: 1

Query Match 56.2%; Score 41; DB 2; Length 692;
 Best Local Similarity 50.0%; Pred. No. 1.9e+02;
 Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 EWTNIWW 8
 Db 345 QWGRDWW 352
 |||||

RESULT 31
 C86314
 Hypothetical protein F2H15.13 - *Arabidopsis thaliana*
 C;Species: *Arabidopsis thaliana* (mouse-ear cress)
 C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
 C;Accession: C86314
 R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alons
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, I
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzial
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A;Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
 A;Reference number: AB6141; MUID:21016719; PMID:11130712
 A;Accession: C86314
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-764 <STO>
 A;Cross-references: GB:AE005172; NID:9665068; PIDN:AAF97270.1; GSPDB:GN00141
 C;Genetics:
 A;Map position: 1
 C;Superfamily: *Arabidopsis* probable serine/threonine-specific protein kinase PRO25; pr

Query Match 55.2%; Score 41; DB 2; Length 764;
Best Local Similarity 83.3%; Pred. No. 2e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 IWWAK 10
|||
Db 397 IWWAK 402

RESULT 32
F72532
hypothetical protein APE2234 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: F72532
R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahara, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; M DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: F72532
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-169 <RAW>
A:Cross-references: DDBJ:AP000063; NID:95105654; PIDN:3AA81246.1; PID:95105934
A:Experimental source: strain K1
C:Genetics:
C:Superfamily: Aeropyrum pernix hypothetical protein APE2234

Query Match 55.5%; Score 40.5; DB 2; Length 169;
Best Local Similarity 71.4%; Pred. No. 53;
Matches 5; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 2 WTNIWW 8
|||
Db 126 WT-LWW 131

RESULT 33
H97039
phage related protein, YonF B. subtilis homolog [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: H97039
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: H97039
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-589 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK79107.1; PID:g15024053; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1134

Query Match 55.5%; Score 40.5; DB 2; Length 589;
Best Local Similarity 55.6%; Pred. No. 1.8e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 1 EW-TNIWW 8
|||
Db 38 EWEKQVWW 46

RESULT 34
B83941
hypothetical protein BH2330 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: B83941
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: B83941
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-139 <STO>
A:Cross-references: GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BA06049.1; GSPDB:G A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2330

Query Match 54.8%; Score 40; DB 2; Length 139;
Best Local Similarity 50.0%; Pred. No. 51;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNIWW 8
|||
Db 69 DWTSLWW 76

RESULT 35
AE2536
hypothetical protein alr7576 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7121
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AE2536
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Itigun Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE2536
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-223 <KUR>
A:Cross-references: GB:AP003602; PIDN:BA077219.1; PID:g17134661; GSPDB:GN00181
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr7576
A:Genome: plasmid

Query Match 54.8%; Score 40; DB 2; Length 223;
Best Local Similarity 55.6%; Pred. No. 82;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNIWW 9
|||
Db 129 EWTDVYLA 137

RESULT 36
A64150
hypothetical protein HI0367 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: A64150
R:Feischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M. Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Ventel A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: A64150
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-303 <TIGR>
A:Cross-references: GB:U32721; GB:L42023; NID:g1573334; PIDN:AAC22025.1; PID:g1573336;

A:Note: best homolog was a hypothetical protein from Escherichia coli

Query Match 54.8%; Score 40; DB 2; Length 303;
Best Local Similarity 66.7%; Pred. No. 1.1e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 TNIWWW 8
DB 124 TGLWWW 129

RESULT 37

D89290
UDP-glucose 4,6-dehydratase (rfbb) homolog - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 16-Jul-1999
C:Accession: D69290
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
-; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: D69290

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
A:Residues: 1-332 <KLE>
A:Cross-references: GB:AE001082; GB:AE000782; NID:G2689405; PIDN:AB90911.1; PID:G265031
C:Superfamily: Escherichia coli UDPglucose 4-epimerase; UDPglucose 4-epimerase homology
F:3-314/Domain: UDPglucose 4-epimerase homology <UDP>

Query Match 54.8%; Score 40; DB 2; Length 332;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

QY 1 EW-TNIWWW 8
DB 305 EWLQNEWWW 314

RESULT 38

B75098
UDP-glucose 4,6-dehydratase (rfbb) PAB0785 - Pyrococcus abyssi (strain Orsay)

C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: B75098
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: B75098
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-333 <KAW>
A:Cross-references: GB:A248286; GB:AL096836; NID:G5458366; PIDN:CAB50087.1; PID:G545859
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: rfbb; PAB0785
C:Superfamily: Escherichia coli UDPglucose 4-epimerase; UDPglucose 4-epimerase homology
F:3-315/Domain: UDPglucose 4-epimerase homology <UDP>

Query Match 54.8%; Score 40; DB 2; Length 333;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

QY 1 EW-TNIWWW 8
DB 306 EWLQNEWWW 315

RESULT 39

A75600

conserved hypothetical protein - Deinococcus radiodurans (strain RL)

C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: A75600
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; I
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans RL.
A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: A75600

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-350 <WHI>

A:Cross-references: GB:AE001862; GB:AE001825; NID:G6460468; PIDN:AAF12252.1; PID:G6460

A:Experimental source: strain RL

C:Genetics:

A:Gene: DRA0063

A:Map position: 2

Query Match 54.8%; Score 40; DB 2; Length 350;
Best Local Similarity 57.1%; Pred. No. 1.3e+02;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 WTNIWWW 8
DB 13 WSAQWWW 19

RESULT 40

S65297

probable membrane protein YPL264c - Yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein P0373
C:Species: Saccharomyces cerevisiae
C:Date: 10-Dec-1994 #sequence_revision 31-May-1996 #text_change 19-Apr-2002
C:Accession: S65297; S65318
R:Dueterhoeft, A.; Floeth, M.; Fritz, M.; Hilbert, H.; Moestl, D.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S65292
A:Accession: S65297

A:Molecule type: DNA

A:Residues: 1-353 <DUE>

A:Cross-references: EMBL:Z73620; NID:G1370544; PID:e246970; PID:G1370545; MIPS:YPL264c
A:Experimental source: strain S288C (AB972)

R:Delius, H.; Hebling, U.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S64967

A:Accession: S65318

A:Molecule type: DNA

A:Residues: 1-353

A:Cross-references: EMBL:Z73620; NID:G1370544; PID:e246970; PID:G1370545; MIPS:YPL264c
A:Experimental source: strain S288C (AB972)

C:Genetics:

A:Cross-references: SGD:S00006185

A:Map position: 16L

C:Keywords: transmembrane protein

F:16-32/Domain: transmembrane #status predicted <TM1>

F:54-70/Domain: transmembrane #status predicted <TM2>

F:96-112/Domain: transmembrane #status predicted <TM3>

F:118-134/Domain: transmembrane #status predicted <TM4>

F:143-159/Domain: transmembrane #status predicted <TM5>

F:189-205/Domain: transmembrane #status predicted <TM6>

F:224-240/Domain: transmembrane #status predicted <TM7>

Query Match 54.8%; Score 40; DB 2; Length 353;
Best Local Similarity 71.4%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 WTNIWWW 8
DB 305 WPNITWM 311

A:Cross-references: GB:AP000001; NID:g3236128; PIDN:BAA29275.1; PID:g3256592
A:Experimental source: strain O73
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH0206

Query Match 54.8%; Score 40; DB 2; Length 544;
Best Local Similarity 44.4%; Pred. No. 2e+02;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 2 WTNLWVWAK 10
|:|||||
216 WSDVYVWAK 224

Db

RESULT 46
JC7520
endo-1,6-alpha-D-mannanase (EC 3.2.1.-) - Bacillus circulans
C:Species: Bacillus circulans
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 30-Jun-2001
C:Accession: JC7520; PC7102
R:Maruyama, Y.; Nakajima, T.
BioSci. Biotechnol. Biochem. 64, 2018-2020, 2000
A:Title: The aman6 gene encoding a yeast mannan backbone degrading 1,6-alpha-D-mannanase
A:Reference number: JC7520; MUID:20506608; PMID:11055417
A:Accession: JC7520
A:Molecule type: DNA
A:Residues: 1-589 <MAN>
A:Cross-references: DDBJ:AB024331
A:Experimental source: strain TN-31
A:Accession: PC7102
A:Molecule type: protein
A:Residues: 36-58;218-253;258-270 <MA2>
C:Comment: This enzyme is a yeast mannan backbone degrading enzyme.
C:Genetics:
A:Gene: aman6
C:Keywords: glycosidase; hydrolase

Query Match 54.8%; Score 40; DB 2; Length 589;
Best Local Similarity 46.7%; Pred. No. 2.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 6; Gaps 1;

QY 1 EWTN-----IWWWA 9
|:|||||
116 DWTNPFNDIMWWA 130

Db

RESULT 47
H87338
sensor histidine kinase / response regulator [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: H87338
R:Nierman, W.C.; Feldblyum, C.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: H87338
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-591 <STO>
A:Cross-references: GB:AE005673; NID:g13421950; PIDN:AAK22708.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0723

Query Match 54.8%; Score 40; DB 2; Length 591;
Best Local Similarity 62.5%; Pred. No. 2.2e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 WTNLWVWAK 9
|:|||||

Db 112 WNNIVVWMS 119

RESULT 48
F95859
probable oligopeptide ABC transporter permease protein Smb20142 [imported] - Sinorhizobium meliloti
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: F95859
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hern
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-bp pSymB megaplasmid from the N2-fixing en
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: F95859
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-307 <KUR>
A:Cross-references: GB:AL591985; PIDN:CAC48542.1; PID:g15140014; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymB
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Huble
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaux
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,
A:Title: The complete genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: Smb20142
A:Genome: plasmid
C:Superfamily: oligopeptide permease protein oppB

Query Match 54.1%; Score 39.5; DB 2; Length 307;
Best Local Similarity 55.6%; Pred. No. 1.3e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 1 EWTNIV-WW 8
|:|||||
69 DWTRTVWVW 77

Db

RESULT 49
A48548
genome polyprotein - coxsackievirus A24 (strain EH24/70)
N:Contains: coat protein 1A; coat protein 1B; coat protein 1C; coat protein 1D; core p
polymerase (EC 2.7.7.48)
C:Species: coxsackievirus A24
C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 05-Jun-1998
R:Supanaranond, K.; Takeda, N.; Yamazaki, S.
Virus Genes 6, 149-158, 1992
A:Title: The complete nucleotide sequence of a variant of Coxsackievirus A24, an agent
A:Reference number: A48548; MUID:92271460; PMID:1317075
A:Accession: A48548
A:Molecule type: genomic RNA
A:Residues: 1-2214 <SUP>
A:Note: sequence extracted from NCBI backbone (NCBI:104525, NCBIP:104526)
C:Superfamily: poliovirus genome polyprotein
C:Keywords: coat protein; core protein; genome-linked protein; nucleotidyltransferase;
F:1-69/Product: coat protein 1A #status predicted <VP4>
F:70-340/Product: coat protein 1B #status predicted <VP2>
F:341-577/Product: coat protein 1C #status predicted <VP3>
F:578-888/Product: coat protein 1D #status predicted <VP1>
F:889-1035/Product: core protein 2A #status predicted <PA2>
F:1036-1132/Product: core protein 2B #status predicted <PB2>
F:1133-1461/Product: core protein 2C #status predicted <PC2>
F:1462-1548/Product: protein 3A #status predicted <PA3>
F:1549-1570/Product: genome-linked protein VPg #status predicted <PB3>
F:1571-1753/Product: proteinase #status predicted <PC3>
F:1754-2214/Product: RNA-directed RNA polymerase #status predicted <PD3>
F:1551/Binding site: phosphoryl-RNA (Tyr) (covalent) #status predicted

A:Gene: Rv2698
C:Superfamily: Mycobacterium tuberculosis hypothetical protein Rv2698

Query Match 53.4%; Score 39; DB 2; Length 161;
Best Local Similarity 57.1%; Pred. No. 81;
Matches 4; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 WTNINWVW 8
DB 18 WVFVWVW 24

RESULT 55
AF3596
phosphatidylcholine synthase (EC 2.7.8.-) [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AF3596
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova, J.; Mazur, M.; Goldsman, E.; Salkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AF3596
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-266 <KUR>
A:Cross-references: GB:AE008918; PIDN:AAU53937.1; PID:G17984882; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEII0695
A:Map position: II
C:Keywords: transferase

Query Match 53.4%; Score 39; DB 2; Length 266;
Best Local Similarity 57.1%; Pred. No. 1.3e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WTNINWVW 8
DB 47 YTAWVWVW 53

RESULT 56
SAVL64
middle surface antigen precursor - woodchuck hepatitis virus (clone 64)
N:Contains: major surface antigen
C:Species: woodchuck hepatitis virus
C>Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 16-Jul-1999
C:Accession: B29498
R:Rtombelle, J.; Moeroey, T.; Trepo, C.; Tiollais, P.; Buendia, M.A.
Gene 50, 207-214, 1986
A:Title: Nucleotide sequence of the woodchuck hepatitis virus surface antigen mRNAs and
A:Reference number: A91568; MUID:87219879; PMID:3582979
A:Accession: B29498
A:Molecule type: mRNA
A:Residues: 1-282 <ETI>
A:Cross-references: GB:M15954; NID:G893289; PIDN:AAA69574.1; PID:G336156
C:Genetics:
A:Gene: pre-S2/S
C:Superfamily: hepatitis B virus surface antigen
C:Keywords: glycoprotein; surface antigen
F:1-60/Domain: signal sequence #status predicted <SIG>
F:61-282/Product: major surface antigen (gene S) #status predicted <MSA>
F:3/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 53.4%; Score 39; DB 1; Length 282;
Best Local Similarity 40.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

QY 2 WTNINWVW 8
DB 80 WTKIPTIAQLNDWVW 94

RESULT 57
F90334
dTDP-Glucose 4,6-dehydratase (rfbB-2) [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C:Accession: F90334
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aweyer, M.J.; Chan
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Feng, X.; Thi-Ngoc, H.P.; Redder,
arrett, R.A.; Ragan, M.A.; Sersen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: F90334
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-350 <KUR>
A:Cross-references: GB:AE006641; NID:G13814979; PIDN:AAK41933.1; GSPDB:GN00155
C:Genetics:
A:Gene: rfbB-2

Query Match 53.4%; Score 39; DB 2; Length 350;
Best Local Similarity 45.5%; Pred. No. 1.8e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

QY 1 EW-TNINWVW 9
DB 318 DWYANNWVW 328

RESULT 58
C88030
Protein F46F5.10 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: C88030
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolo
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.el
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; an
A:Accession: C88030
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-369 <STO>
A:Cross-references: GB:chr_II; PIDN:AC78188.1; PID:G3886037; GSPDB:GN00020; CESP:F46F5.
C:Genetics:
A:Gene: F46F5.10
A:Map position: 2
C:Superfamily: Caenorhabditis elegans hypothetical protein F22E5.1

Query Match 53.4%; Score 39; DB 2; Length 369;
Best Local Similarity 57.1%; Pred. No. 1.9e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 EWTNINWVW 7
DB 202 EYSNVWVW 208

RESULT 59
JCS461
cellulase (EC 3.2.1.4) precursor - imperfect fungus (Humicola grisea)
N:Alternate names: endo-1,4-beta-glucanase; endoglucanase
C:Species: Humicola grisea var. thermoides
C>Date: 11-Sep-1998 #sequence_revision 11-Sep-1998 #text_change 16-Jun-2000
C:Accession: JCS461
R:Takahama, S.; Nakamura, A.; Masaki, H.; Uozumi, T.
Biosci. Biotechnol. Biochem. 61, 245-250, 1997
A:Title: Cloning, sequencing, and expression of a thermostable cellulase gene of Humico
A:Reference number: JCS461; MUID:97212020; PMID:9058960

A:Accession: J05461
A:Molecule type: DNA
A:Residues: 1-388 <TAK>
A:Cross-references: DDBJ:D84470; NID:G1304101; PIDN:BAAL2676.1; PID:G1304102
A:Experimental source: strain IFO9854
C:Genetics:
A:Gene: egl2
A:Introns: 120/3; 369/1
C:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellulose
A:Pathway: cellulose degradation
C:Superfamily: Trichoderma cellulase III; fungal cellulose-binding domain homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-388/Product: cellulase #status predicted <MAT>
F:21-52/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 53.4%; Score 39; DB 1; Length 388;
Best Local Similarity 62.5%; Pred. No. 2e+02; Mismatches 3; Indels 0; Gaps 0;
Matches 5; Conservative 0

QY 2 WTNWMA 9
DB 349 WTGALWMA 356

RESULT 60
S43920
cellulase (EC 3.2.1.4) precursor - imperfect fungus (Humicola insolens)
N:Alternate names: endo-1,4-beta-glucanase; endoglucanase
C:Species: Humicola insolens
C:Date: 11-Sep-1998 #sequence_revision 11-Sep-1998 #text_change 18-Jun-1999
C:Accession: S43920
R:Dalboge, H.; Heldt-Hansen, H.P.
Mol. Genet. 243, 253-260, 1994
A:Title: A novel method for efficient expression cloning of fungal enzyme genes.
A:Reference number: S43919; MUID:94247364; PMID:8190078
A:Accession: S43920
A:Molecule type: mRNA
A:Residues: 1-388 <DAL>
A:Cross-references: EMBL:X76046; NID:G505194; PIDN:CAA53631.1; PID:G505195
C:Genetics:
A:Gene: CMC3
A:Function:
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellulose
A:Pathway: cellulose degradation
C:Superfamily: Trichoderma cellulase III; fungal cellulose-binding domain homology
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-388/Product: cellulase #status predicted <MAT>
F:21-52/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 53.4%; Score 39; DB 1; Length 388;
Best Local Similarity 62.5%; Pred. No. 2e+02; Mismatches 3; Indels 0; Gaps 0;
Matches 5; Conservative 0

QY 2 WTNWMA 9
DB 349 WTGALWMA 356

RESULT 61
C84518
Hypothetical protein At2g14530 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: C84518
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.J.; Town, C.D.; Fujii, C.Y.; M.; Koo, R.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.; Eures, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:2C081487; PMID:10617197

A:Accession: C84518
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-412 <STO>
A:Cross-references: GB:AE002093; NID:G4263820; PIDN:AAI5463.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g14530
A:Map position: 2

Query Match 53.4%; Score 39; DB 2; Length 412;
Best Local Similarity 71.4%; Pred. No. 2.1e+02; Mismatches 2; Indels 0; Gaps 0;
Matches 5; Conservative 0

QY 3 TNNWMA 9
DB 237 TGHWWMA 243

RESULT 62
S53075
probable membrane protein YMR253c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YMR253c
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
C:Accession: S53075
R:Hunt, S.; Bowman, S.
submitted to the EMBL Data Library, March 1995
A:Reference number: S53069
A:Accession: S53075
A:Molecule type: DNA
A:Residues: 1-414 <HUN>
A:Cross-references: EMBL:Z48639; NID:G732924; PID:G732931; GSPDB:GN00013; MIPS:YMR253C
C:Genetics:
A:Gene: MIPS:YMR253C
A:Cross-references: SGD:S0004866
A:Map position: 13R
C:Keywords: transmembrane protein
F:59-85/Domain: transmembrane #status predicted <TM1>
F:109-125/Domain: transmembrane #status predicted <TM2>
F:174-190/Domain: transmembrane #status predicted <TM3>
F:198-214/Domain: transmembrane #status predicted <TM4>
F:246-262/Domain: transmembrane #status predicted <TM5>
F:272-288/Domain: transmembrane #status predicted <TM6>
F:308-324/Domain: transmembrane #status predicted <TM7>
F:364-380/Domain: transmembrane #status predicted <TM8>

Query Match 53.4%; Score 39; DB 2; Length 414;
Best Local Similarity 71.4%; Pred. No. 2.1e+02; Mismatches 2; Indels 0; Gaps 0;
Matches 5; Conservative 0

QY 2 WTNWMA 8
DB 358 WPNWMA 364

RESULT 63
SAVLC
large surface antigen - woodchuck hepatitis virus (clone 1)
N:Contains: major surface antigen; middle surface antigen
C:Species: woodchuck hepatitis virus
C:Date: 14-Nov-1983 #sequence_revision 14-Nov-1983 #text_change 16-Jul-1999
C:Accession: A03707
R:Galibert, F.; Chen, T.N.; Mandart, E.
J. Virol. 41, 51-65, 1982
A:Title: Nucleotide sequence of a cloned woodchuck hepatitis virus genome: comparison
A:Reference number: A92886; MUID:82216969; PMID:7086958
A:Accession: A03707
A:Molecule type: DNA
A:Residues: 1-426 <GML>
A:Cross-references: GB:J02442; NID:G336126; PIDN:AAA46760.1; PID:G336128
C:Genetics:
A:Gene: pre-S1/pre-S2/S
C:Superfamily: hepatitis B virus surface antigen

C;Keywords: glycoprotein; surface antigen
F:145-426/Product: middle surface antigen (gene pre-S2/S) #status predicted <DSA>
F:205-426/Product: major surface antigen (gene S) #status predicted <MSA>
F:32,147,346/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 53.4%; Score 39; DB 1; Length 426;
Best Local Similarity 40.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

QY 2 WTNI-----WWW 8
|||
224 WTKILTTIAQLNDWWW 238

Db

RESULT 64
SAVL5
large surface antigen - ground squirrel hepatitis virus
N;Contains: major surface antigen; middle surface antigen
C;Species: ground squirrel hepatitis virus
C;Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 15-Nov-1996
C;Accession: A03709
R;Seeger, C.; Ganem, D.; Varmus, H.E.
J. Virol. 51, 367-375, 1984
A;Title: Nucleotide sequence of an infectious molecularly cloned genome of ground squirrel hepatitis virus
A;Reference number: A93000; MUID:84267998; PMID:6086950
A;Accession: A03709
A;Molecule type: DNA
A;Residues: 1-428 <SEE>
A;Cross-references: GB:K02715
C;Genetics:
A;Gene: pre-S1/pre-S2/S
C;Superfamily: hepatitis B virus surface antigen
C;Keywords: glycoprotein; surface antigen
F:147-428/Product: middle surface antigen (gene pre-S2/S) #status predicted <DSA>
F:207-428/Product: major surface antigen (gene S) #status predicted <MSA>
F:32,147,346/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 53.4%; Score 39; DB 1; Length 428;
Best Local Similarity 40.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

QY 2 WTNI-----WWW 8
|||
226 WTKILTTIAQLNDWWW 240

Db

RESULT 65
SAVL7
large surface antigen - woodchuck hepatitis virus (clone 7)
N;Contains: major surface antigen; middle surface antigen
C;Species: woodchuck hepatitis virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C;Accession: D29969
R;Cohen, J.I.; Miller, R.H.; Rosenblum, B.; Denniston, K.; Gerin, J.L.; Purcell, R.H.
Virology 162, 12-20, 1988
A;Title: Sequence comparison of woodchuck hepatitis virus replicative forms shows conserved regions
A;Reference number: A94368; MUID:88101359; PMID:3336938
A;Accession: D29969
A;Molecule type: DNA
A;Residues: 1-431 <COH>
A;Cross-references: GB:M18752; NID:G336136; PIDN:AAA46766.1; PID:G336137
C;Genetics:
A;Gene: pre-S1/pre-S2/S
C;Superfamily: hepatitis B virus surface antigen
C;Keywords: glycoprotein; surface antigen
F:150-431/Product: middle surface antigen (gene pre-S2/S) #status predicted <DSA>
F:210-431/Product: major surface antigen (gene S) #status predicted <MSA>
F:32,94,152/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 53.4%; Score 39; DB 1; Length 431;
Best Local Similarity 40.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

QY 2 WTNI-----WWW 8
|||
229 WTKILTTIAQLNDWWW 243

Db

RESULT 66
SAVLW8
large surface antigen - woodchuck hepatitis virus (clone 8)
N;Contains: major surface antigen; middle surface antigen
C;Species: woodchuck hepatitis virus
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 12-Jun-1998
C;Accession: B32397
R;Girones, R.; Cote, P.J.; Hornbuckle, W.E.; Tennant, B.C.; Gerin, J.L.; Purcell, R.H.;
Proc. Natl. Acad. Sci. U.S.A. 86, 1846-1849, 1989
A;Title: Complete nucleotide sequence of a molecular clone of woodchuck hepatitis virus
A;Reference number: A94222; MUID:89184524; PMID:2928306
A;Accession: B32397
A;Molecule type: DNA
A;Residues: 1-431 <GIR>
A;Cross-references: GB:J04514
C;Genetics:
A;Gene: pre-S1/pre-S2/S
C;Superfamily: hepatitis B virus surface antigen
C;Keywords: glycoprotein; surface antigen
F:150-431/Product: middle surface antigen (gene pre-S2/S) #status predicted <DSA>
F:210-431/Product: major surface antigen (gene S) #status predicted <MSA>
F:32,94,152/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 53.4%; Score 39; DB 1; Length 431;
Best Local Similarity 40.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

QY 2 WTNI-----WWW 8
|||
229 WTKILTTIAQLNDWWW 243

Db

RESULT 67
SAVL59
large surface antigen - woodchuck hepatitis virus (clone 59)
N;Contains: major surface antigen; middle surface antigen
C;Species: woodchuck hepatitis virus
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C;Accession: H29969
R;Cohen, J.I.; Miller, R.H.; Rosenblum, B.; Denniston, K.; Gerin, J.L.; Purcell, R.H.
Virology 162, 12-20, 1988
A;Title: Sequence comparison of woodchuck hepatitis virus replicative forms shows conserved regions
A;Reference number: A94368; MUID:88101359; PMID:3336938
A;Accession: H29969
A;Molecule type: DNA
A;Residues: 1-431 <COH>
A;Cross-references: GB:M19183; NID:G336141; PIDN:AAA46762.1; PID:G336142
C;Genetics:
A;Gene: pre-S1/pre-S2/S
C;Superfamily: hepatitis B virus surface antigen
C;Keywords: glycoprotein; surface antigen
F:150-431/Product: middle surface antigen (gene pre-S2/S) #status predicted <DSA>
F:210-431/Product: major surface antigen (gene S) #status predicted <MSA>
F:32,94,152/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 53.4%; Score 39; DB 1; Length 431;
Best Local Similarity 40.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

QY 2 WTNI-----WWW 8
|||
229 WTKILTTIAQLNDWWW 243

Db

RESULT 68
596837
unknown protein T21F11.12 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: E96837
R:Theologas, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso-
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; XUID:21016719; PMID:11130712
A:Accession: E96837
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-448 <STO>
A:Cross-references: GB:AB005173; NID:g6730729; PIDN:AAF27119.1; GSPDB:GN00141
C:Genetics:
A:Gene: T21F11.12
A:Map position: 1

Query Match 53.4%; Score 39; DB 2; Length 448;
Best Local Similarity 44.4%; Pred. No. 2.3e+02;
Matches 4; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 WTNWVWAK 10
| : : : :
Db 199 WSKLGWGWK 207

RESULT 69
G90500
glycolate oxidase glcd subunit (glcd) [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 15-Jun-2001
C:Accession: G90500
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Avayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, H.
arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: G90500
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-467 <KOR>
A:Cross-references: GB:AE006641; NID:gl3816591; PIDN:AAK43262.1; GSPDB:GN00155
C:Genetics:
A:Gene: glcd
C:Superfamily: glycolate oxidase chain glcd

Query Match 53.4%; Score 39; DB 2; Length 467;
Best Local Similarity 50.0%; Pred. No. 2.3e+02;
Matches 5; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 EWTNWWAK 10
| : : : :
Db 300 EENRWVWGR 309

RESULT 70
G69670
glycine betaine transporter opuD - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C:Accession: G69670
R:Kunst, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte-
C.; Bron, S.; Brouillet, S.; Bruchi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch-
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler-
iech, J.; Harwood, C.R.; Hentaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlo
A:Authors: Schleich, S.; Schroter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Ser
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terstra, P.; Tognoni, A.; Tosato, V.; Uchiyam
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida,
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: G69670
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-512 <KUN>
A:Cross-references: GB:Z99119; GB:AL009126; NID:g2635411; PIDN:CAB14985.1; PID:g263549
A:Experimental source: strain 168
C:Genetics:
A:Gene: opuD
C:Superfamily: Escherichia coli probable carnitine transport protein

Query Match 53.4%; Score 39; DB 2; Length 512;
Best Local Similarity 42.9%; Pred. No. 2.6e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 6; Gaps 1;

Qy 1 EWTNWWAK 8
| : : : :
Db 307 EWTNWWAK 320

RESULT 71
I38607
p53-binding protein 2 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 09-Mar-1996 #sequence_revision 09-Mar-1996 #text_change 20-Sep-1999
C:Accession: I38607
R:Iwabuchi, K.; Bartel, P.L.; Li, B.; Marraccino, R.; Fields, S.
Proc. Natl. Acad. Sci. U.S.A. 91, 6098-6102, 1994
A:Title: Two cellular proteins that bind to wild-type but not mutant p53.
A:Reference number: I38604; MUID:94285584; PMID:8016121
A:Accession: I38607
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-529 <RES>
A:Cross-references: EMBL:U09582; NID:g430379; PIDN:AAA21597.1; PID:g493080
C:Superfamily: unassigned ankryrin repeat proteins; ankryrin repeat homology; SH3 homolo
F;392-424/Domain: ankryrin repeat homology <AN08>
F;465-515/Domain: SH3 homology <SH3>

Query Match 53.4%; Score 39; DB 2; Length 529;
Best Local Similarity 80.0%; Pred. No. 2.7e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 6 WWWWAK 10
| : : : :
Db 498 WWWWAK 502

RESULT 72
T41341
probable serine-threonine-protein kinase - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 19-Jul-2002
C:Accession: T41341
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL Data Library, January 1999
A:Reference number: Z21970
A:Accession: T41341
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-624 <LYN>
A:Cross-references: EMBL:AL035076; PIDN:CAA22652.1; GSPDB:GN00066; SPDB:SPCC417.06C
A:Experimental source: strain 972h-; cosmid c417
C:Genetics:

A:Gene: SPDB:SPCC417.06C
A:Map position: -
A:Introns: 203/3; 353/3
C:Superfamily: protein kinase DBF2; protein kinase homology

Query Match 53.4%; Score 39; DB 2; Length 624;
Best Local Similarity 57.1%; Pred. No. 3.1e+02;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 WTNLWVW 8
|||
Db 406 WTNLYYW 412

RESULT 73
G69371
acetyl-CoA synthetase (acs-4) homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
C:Accession: G69371
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.P.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
-; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiaich, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: G69371
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-642 <KLE>
A:Cross-references: GB:AB001037; GB:AE000782; NID:G2689360; PIDN:AAB90268.1; PTD:G264962
C:Superfamily: acetate-CoA ligase; acetate-CoA ligase homology
F:134-613/Domain: acetate-CoA ligase homology <ACL>

Query Match 53.4%; Score 39; DB 2; Length 642;
Best Local Similarity 80.0%; Pred. No. 3.2e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 WNWAK 10
|||
Db 45 WWSK 49

RESULT 74
S66129
disintegrin (EC 3.4.24.-) precursor - bovine
C:Species: Bos primigenius taurus (cattle)
C>Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 05-Nov-1999
C:Accession: S66129; S32205
R:Glynn, P.; Howard, L.
submitted to the EMBL Data Library, October 1995
A:Description: identification of a mammalian member of the metalloproteinase/disintegrin
A:Reference number: S66129
A:Accession: S66129
A:Molecule type: mRNA
A:Residues: 1-748 <GLY>
A:Cross-references: EMBL:Z21961; NID:G1044810; PIDN:CAA79973.1; PTD:G1044811
A:Note: this is a revision to the sequence from reference S32205
R:Glynn, P.; Howard, L.
submitted to the EMBL Data Library, March 1993
A:Description: identification of a mammalian member of the metalloproteinase/disintegrin
A:Reference number: S32205
A:Accession: S32205
A:Molecule type: mRNA
A:Residues: 1-113 'LAM', 117 'LLLMEDLKDSFRLMVAR', 135-171, 'R', 173-652, 'L' <GLW>
A:Cross-references: EMBL:Z21961
A:Note: this sequence has been revised in reference S66129
C:Superfamily: disintegrin homology
C:Keywords: hydrolase; metalloproteinase; zinc
F:1-13/Domain: signal sequence #status predicted <SIG>
F:14-748/Product: metalloproteinase #status predicted <MAT>

F:456-546/Domain: disintegrin homology <DIS>
F:383,387,393/Binding site: zinc, catalytic (His) #status predicted
F:384/Active site: Glu #status predicted

Query Match 53.4%; Score 39; DB 2; Length 748;
Best Local Similarity 55.6%; Pred. No. 3.8e+02;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EWTNIWWWA 9
|||
Db 672 EWIVAYWWA 680

RESULT 75
JC7537
beta-N-acetylglucosaminidase - Alteromonas sp. (strain O-7)
C:Species: Alteromonas sp. (strain O-7)
C>Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 01-Mar-2002
C:Accession: JC7537; PC7108
R:Tsujibo, R.; Miyamoto, J.; Kondo, N.; Miyamoto, K.; Baba, N.; Inamori, Y.
Biosci. Biotechnol. Biochem. 64, 2512-2516, 2000
A:Title: Molecular cloning of the gene encoding an outer-membrane-associated beta-N-ac
A:Reference number: JC7537; MUID: 21036923; PMID:11193430
A:Accession: JC7537
A:Molecule type: DNA
A:Residues: 1-863 <TSU>
A:Cross-references: DBJ:AB042622
A:Experimental source: strain O-7
A:Accession: PC7108
A:Molecule type: protein
A:Residues: 336-356; 464-471; 520-531; 548-557; 609-617; 716-726 <TS2>
C:Comment: This enzyme, an outer-membrane-associated lipoprotein, which belongs to the
glucosamine.
C:Genetics:
A:Gene: GlcNAcaseA
C:Superfamily: beta-hexosaminidase
C:Keywords: lipoprotein

Query Match 53.4%; Score 39; DB 2; Length 863;
Best Local Similarity 71.4%; Pred. No. 4.3e+02;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 WTNLWVW 8
|||
Db 622 WDNLWVW 628

Search completed: June 9, 2004, 18:11:29
Job time : 31 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2004, 18:02:56 ; Search time 11 Seconds
(without alignments)

47.337 Million cell updates/sec

Title: US-10-726-148A-15_COPY_428_437

Perfect score: 73

Sequence: 1 EWTNIWWAK 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARYES

Result No.	Score	Query Match %	Length	DB ID	Description
1	44	60.3	14	1 LPW RHINE	P18854 rhizobium m
2	43.5	59.6	2236	1 PYR1 DROME	P05990 drosophila
3	43	58.9	572	1 FYDT USTMA	Q92459 ustilago ma
4	43	58.9	688	1 YAOA SCHPO	Q10109 schizosacch
5	43	58.9	938	1 CDG8 HUMAN	Q95477 homo sapien
6	42.5	58.2	473	1 YID9 YEAST	P40333 saccharomyc
7	42.5	58.2	720	1 YD71 SCHPO	Q10326 schizosacch
8	41	56.2	337	1 DNJ2 CAEEL	Q17433 caenorhabdi
9	41	56.2	372	1 SUM1 MOUSE	Q8r0f3 mus musculu
10	41	56.2	374	1 SUM1 HUMAN	Q8nbk3 homo sapien
11	41	56.2	493	1 YLD2 CAEEL	Q03567 caenorhabdi
12	41	56.2	562	1 SYK AERPE	Q9yft9 aeropyrum p
13	40	54.8	303	1 Y367 HAEIN	Q57065 haemophilus
14	39.5	54.1	2214	1 POLG CXA24	P36290 c genome po
15	39	53.4	282	1 VMSA WHV6	P11293 woodchuck h
16	39	53.4	350	1 RFBB RHISN	P55462 rhizobium s
17	39	53.4	388	1 GUN3 HUMIN	Q12624 humicola in
18	39	53.4	414	1 YMS7 YEAST	Q04835 saccharomyc
19	39	53.4	426	1 VMSA WHV1	P03143 woodchuck h
20	39	53.4	428	1 VMSA HPBVS	P03144 ground squi
21	39	53.4	431	1 VMSA WHV59	P12910 woodchuck h
22	39	53.4	431	1 VMSA WHV7	P12909 woodchuck h
23	39	53.4	431	1 VMSA WHV81	P17400 woodchuck h
24	39	53.4	451	1 PTK6 MOUSE	Q64434 mus musculu
25	39	53.4	501	1 NUOM BUCAP	Q8k9x6 buchnera ap
26	39	53.4	512	1 OPUD BACSU	P54417 bacillus su
27	39	53.4	558	1 VNLI DROME	Q9nfp1 drosophila
28	39	53.4	748	1 AD10 BOVIN	Q10741 bos taurus
29	39	53.4	842	1 ABCG HUMAN	Q9np58 homo sapien
30	39	53.4	904	1 YNS3 YEAST	P42842 saccharomyc
31	39	53.4	912	1 MGR4 HUMAN	Q14833 homo sapien
32	39	53.4	912	1 MGR4 RAT	P31423 rattus norv
33	39	53.4	1087	1 ASP1 MOUSE	Q62415 mus musculu

Q8cg79 mus musculu	1088	1	ASP2 MOUSE	1088	39	53.4	34
Q96kq4 homo sapien	1090	1	ASP1 HUMAN	1090	39	53.4	35
Q13625 homo sapien	1128	1	ASP2 HUMAN	1128	39	53.4	36
P02954 rhodobacter	281	1	RCEL RHOSH	281	38.5	52.7	37
Q9ycy3 aeropyrum p	572	1	SYM AERPE	572	38.5	52.7	38
P35867 corynebacte	794	1	Y966 CORGL	794	39	52.7	39
Q98mn3 rhizobium l	268	1	PCS RHILLO	268	38	52.1	40
P11409 proteus vul	336	1	MTF2 PROVU	336	38	52.1	41
Q9lfw3 arabidopsis	431	1	CBL4 ARATH	431	38	52.1	42
Q8xao3 escherichia	504	1	CAIT ECO57	504	38	52.1	43
P59333 escherichia	504	1	CAIT ECO57	504	38	52.1	44
P15533 escherichia	504	1	CAIT ECO57	504	38	52.1	45
P59334 proteus sp.	504	1	CAIT PROSL	504	38	52.1	46
P59335 shigella fl	504	1	CAIT SHFL	504	38	52.1	47
C8z911 salmonella	505	1	CAIT SALT1	505	38	52.1	48
Q8zrx1 salmonella	505	1	CAIT SALT1	505	38	52.1	49
P47342 mycoplasma	650	1	Y096 MYCGE	650	38	52.1	50
Q9a6r5 caulobacter	663	1	OPGH CAUCR	663	38	52.1	51
P17447 escherichia	677	1	BETT ECO57	677	38	52.1	52
Q10768 mycobacteri	765	1	TREY MYCTU	765	38	52.1	53
P03376 human immun	856	1	ENV HVLPV	856	38	52.1	54
P05881 human immun	856	1	ENV HVLPV	856	38	52.1	55
P04579 human immun	865	1	ENV HVLPV	865	38	52.1	56
Q9kiy0 haemophilus	993	1	HGBC HAEIN	993	38	52.1	57
Q9kiy2 haemophilus	1013	1	HGBC HAEIN	1013	38	52.1	58
Q9m9w8 arabidopsis	1039	1	PD2 ARATH	1039	38	52.1	59
P07210 h genome po	2164	1	POLG HRV89	2164	37.5	51.4	60
P22100 vibrio para	41	1	LPW VIBPA	41	37	50.7	61
P72851 synecocyst	78	1	KL28 SYNI3	78	37	50.7	62
Q53229 rhodobacter	165	1	YRH1 RHOSH	165	37	50.7	63
Q18864 caenorhabdi	277	1	SUR4 CAEEL	277	37	50.7	64
Q8wuf5 homo sapien	407	1	IASF HUMAN	407	37	50.7	65
Q13882 homo sapien	451	1	PTK6 HUMAN	451	37	50.7	66
Q9sr77 arabidopsis	452	1	CBL1 ARATH	452	37	50.7	67
Q94kt8 arabidopsis	456	1	COBR ARATH	456	37	50.7	68
Q9qzm2 mus musculu	459	1	DPG2 MOUSE	459	37	50.7	69
O79410 scyflorhinu	460	1	NU4M SCYCA	460	37	50.7	70
Q9ul66 homo sapien	477	1	KCS2 HUMAN	477	37	50.7	71
Q35174 mus musculu	477	1	KCS2 MOUSE	477	37	50.7	72
Q9er26 rattus norv	477	1	KCS2 RAT	477	37	50.7	73
Q9uhni homo sapien	485	1	DPG2 HUMAN	485	37	50.7	74
Q9bq31 homo sapien	491	1	KCS3 HUMAN	491	37	50.7	75
Q9tt17 oryctolagus	491	1	KCS3 RABIT	491	37	50.7	76
Q88759 rattus norv	491	1	KCS3 RAT	491	37	50.7	77
P09700 human cytom	500	1	US24 HOMVA	500	37	50.7	78
Q8tdn2 homo sapien	545	1	KCV2 HUMAN	545	37	50.7	79
Q93852 candida alb	557	1	ALO CANAL	557	37	50.7	80
Q9eqd0 mus musculu	577	1	FZD5 MOUSE	577	37	50.7	81
P56161 anophelae s	664	1	ACES ANOST	664	37	50.7	82
Q44315 arthrobacte	775	1	TREY ARTSQ	775	37	50.7	83
Q96q91 homo sapien	983	1	B3A4 HUMAN	983	37	50.7	84
P17326 artemia san	996	1	AT1A ARTSF	996	37	50.7	85
Q9p2n4 homo sapien	1395	1	AT59 HUMAN	1395	37	50.7	86
C56075 p genome po	3099	1	POLG PEMVM	3099	37	50.7	87
P25818 arabidopsis	251	1	TI11 ARATH	251	36.5	50.0	88
Q41963 arabidopsis	253	1	TI12 ARATH	253	36.5	50.0	89
Q85786 synecococc	263	1	THID SYN7	263	36.5	50.0	90
Q83338 treponema p	59	1	Y318 TREPA	59	36	49.3	91
Q86243 haemophilus	59	1	YF6A HAEIN	59	36	49.3	92
P54942 bacillus su	132	1	YXEC BACSU	132	36	49.3	93
P51751 thiodospiril	255	1	RCEM RHOPH	255	36	49.3	94
P56403 rattus norv	303	1	AQP7 RAT	303	36	49.3	95
O54794 mus musculu	309	1	AQP7 MOUSE	309	36	49.3	96
P35162 bacillus su	352	1	RESC BACSU	352	36	49.3	97
Q80400 humulus lup	394	1	VPS HUMLU	394	36	49.3	98
Q43092 schizosacch	409	1	OX12 SCHPO	409	36	49.3	99
P58599 ralstonia s	424	1	GUN1 RALSO	424	36	49.3	100
P17974 ralstonia s	426	1	GUN1 RALSO	426	36	49.3	101
Q91085 celeagris g	457	1	VIPR MELGA	457	36	49.3	102
Q9zm4 salmo salar	460	1	NU4M SALSA	460	36	49.3	103
Q9w697 xenopus lae	463	1	DPG2 XENLA	463	36	49.3	104
P16750 human cytom	466	1	UL74 HCMVA	466	36	49.3	105
P42403 bacillus su	477	1	BGL2 BACSU	477	36	49.3	106

107	36	49.3	477	1	YBHI_ECOLI	P75763	escherichia	180	34.5	47.3	2208	1	POLH_POLLM	P03300	p genome po
108	36	49.3	547	1	ILVG_MYCTU	Q50613	mycobacteri	181	34.5	47.3	2209	1	POLG_POLLS	P03301	p genome po
109	36	49.3	550	1	PUTX_EWENI	P18696	emericella	182	34	46.6	101	1	PHS_FALSO	O8X038	raistonia s
110	36	49.3	746	1	MEPA_HUMAN	Q16819	homo sapien	183	34	46.6	102	1	YPTH_BOVIN	P09894	bos taurus
111	36	49.3	747	1	MEPA_MOUSE	P28825	mus musculus	184	34	46.6	130	1	YFER_ECOLI	P76549	escherichia
112	36	49.3	748	1	MEPA_RAT	P64230	rattus norv	185	34	46.6	136	1	YCV4 YEAST	P25640	saccharomyc
113	36	49.3	749	1	CATA_LEGEM	Q9WXB9	legionella	186	34	46.6	138	1	YA91_MYCPN	P75602	mycoplasma
114	36	49.3	768	1	ENV_SIVAI	P27757	sinian immu	187	34	46.6	180	1	FATE_BOVIN	Q951a0	bos taurus
115	36	49.3	797	1	VG48_HSVSA	Q01033	herpesvirus	188	34	46.6	187	1	Y893_HAEIN	P44923	haemophilus
116	36	49.3	868	1	ENV_HVIC4	P05879	human immu	189	34	46.6	215	1	G25L_CANFA	P27869	canis fami
117	36	49.3	882	1	CTIB_FUSSO	P52959	fusarium so	190	34	46.6	223	1	C79A_BOVIN	P40293	bos taurus
118	36	49.3	909	1	CTIA_FUSSO	P52958	fusarium so	191	34	46.6	238	1	YUHA_ECOLI	P39372	escherichia
119	36	49.3	964	1	YQKA_SCHPO	O74522	schizosacch	192	34	46.6	249	1	CDSA_ECOLI	P06466	e phosphati
120	36	49.3	1291	1	SETB_HUMAN	Q15047	homo sapien	193	34	46.6	252	1	RCEL_ACICY	O66137	acidiphiliu
121	36	49.3	1307	1	SETB_MOUSE	O89974	mus musculus	194	34	46.6	252	1	Y132_BUCAP	O8K402	buchnera ap
122	36	49.3	1532	1	GDE_HUMAN	P35573	h glycogen	195	34	46.6	254	1	RCEL_ACIOR	O66141	acidiphiliu
123	36	49.3	1555	1	GDE_RABIT	P35574	o glycogen	196	34	46.6	255	1	RCEL_ACIUM	O66139	acidiphiliu
124	36	49.3	2895	1	HYD_DROME	P51592	drosophila	197	34	46.6	255	1	RCEL_RHOPI	P51750	rhodospiril
125	35.5	48.6	718	1	PNT2_DROME	P51023	drosophila	198	34	46.6	265	1	DDHC_RHOSU	Q89291	rhodovulum
126	35.5	48.6	2184	1	POLG_ECO1.F	O91734	e genome po	199	34	46.6	273	1	RCEL_RHOVI	P06009	rhodopseudo
127	35	47.9	100	1	Y05E_BPT4	P39260	bacterioph	200	34	46.6	278	1	RCEL_CHRVI	P51762	chromatium
128	35	47.9	219	1	VATA_STRAU	P26839	staphylococ	201	34	46.6	278	1	RCEL_RHOGE	P51760	rhodocyclus
129	35	47.9	256	1	BIQH_ECOLI	P13001	escherichia	202	34	46.6	279	1	ELOI_HUMAN	Q9B460	homo sapien
130	35	47.9	256	1	Y139_BUCAI	P57239	buchnera ap	203	34	46.6	279	1	ELOI_MOUSE	Q9J1J5	mus musculus
131	35	47.9	268	1	YCT3_HAEIN	P44150	haemophilus	204	34	46.6	279	1	YP69_MYCLE	Q49757	mycobacteri
132	35	47.9	276	1	RLPG_ECOLI	P09391	escherichia	205	34	46.6	281	1	RCEL_RHOCA	P19057	rhodobacter
133	35	47.9	307	1	REP_BACSP	P36229	bacillus sp	206	34	46.6	337	1	VINT_3PP2	P36932	bacterioph
134	35	47.9	317	1	Y302_MYCPN	P75357	mycoplasma	207	34	46.6	340	1	CHI6_POPTR	P16579	populus tri
135	35	47.9	332	1	GLYG_MOUSE	Q9R062	mus musculus	208	34	46.6	340	1	YCEG_ECOLI	P28106	escherichia
136	35	47.9	332	1	GLYG_RAT	O08730	rattus norv	209	34	46.6	344	1	HENZ_AGRIS	Q8U917	agrobacteri
137	35	47.9	339	1	REP_BACAM	P13963	bagittus am	210	34	46.6	366	1	HTPX_AERPE	Q9Y467	aeropyrum p
138	35	47.9	367	1	YMY7 YEAST	Q03151	saccharomyc	211	34	46.6	374	1	FES_ECOLI	P13039	escherichia
139	35	47.9	379	1	CYDB_ECOLI	P11027	escherichia	212	34	46.6	375	1	LDB1_MOUSE	P70662	mus musculus
140	35	47.9	382	1	FATB_CINCA	Q39473	cinnamomum	213	34	46.6	375	1	LDB1_XENLA	P70060	xenopus lae
141	35	47.9	382	1	FATB_UMBCA	Q44635	umbeulicari	214	34	46.6	391	1	CHSY_DIACA	P48389	dianthus ca
142	35	47.9	397	1	YBP2 YEAST	P38226	saccharomyc	215	34	46.6	391	1	CHSY_DIAMO	Q91KP7	dianthus mo
143	35	47.9	399	1	CK05_MOUSE	Q9Y666	mus musculus	216	34	46.6	396	1	YD18 YEAST	Q12185	saccharomyc
144	35	47.9	447	1	Y341_CHXND	Q9P8J6	chlamydia m	217	34	46.6	408	1	METK_DROME	P04320	drosophila
145	35	47.9	448	1	Y347_CHLEN	O9Z8J6	chlamydia p	218	34	46.6	409	1	MDP1_PIG	P24122	sus scrofa
146	35	47.9	451	1	Y069_CHLEF	O84072	chlamydia t	219	34	46.6	410	1	MDP1_RABIT	P31429	oryctolagus
147	35	47.9	461	1	NU4M_XENLA	P03912	xenopus lae	220	34	46.6	424	1	VP3_BPHK1	P49859	bacterioph
148	35	47.9	493	1	VPE2_PHAUV	O24326	phaseolus v	221	34	46.6	429	1	AG45_MYCLE	P46838	mycobacteri
149	35	47.9	510	1	HEX6_RICCO	Q27423	ricinus com	222	34	46.6	433	1	ANM2_HUMAN	P55345	homo sapien
150	35	47.9	522	1	STP1_ARATH	P23586	arabidopsis	223	34	46.6	443	1	UVSH_EWENI	O02398	emericella
151	35	47.9	544	1	AD10_RAT	Q10743	rattus norv	224	34	46.6	448	1	ANM2_MOUSE	Q91144	mus musculus
152	35	47.9	553	1	SUTI_SCHPO	O14091	schizosacch	225	34	46.6	460	1	NU4M_CARAU	O78687	carassius a
153	35	47.9	564	1	YUES_CAEEL	P90859	caenorhabdi	226	34	46.6	462	1	ACHO_CARAU	P13508	carassius a
154	35	47.9	590	1	IRA2_HUMAN	O43187	homo sapien	227	34	46.6	464	1	NU4M_PARLI	P12775	paracentrot
155	35	47.9	648	1	GRT1_SCHPO	Q9C469	schizosacch	228	34	46.6	466	1	ACHP_CARAU	P18257	carassius a
156	35	47.9	748	1	AD10_HUMAN	O14672	homo sapien	229	34	46.6	471	1	KIW7_HUMAN	O76014	homo sapien
157	35	47.9	749	1	AD10_MOUSE	O35598	mus musculus	230	34	46.6	473	1	Y309_SYNV3	O55914	synecocyst
158	35	47.9	749	1	AD10_XENLA	O81Y1	xenopus lae	231	34	46.6	482	1	VGLY_TACV7	O55914	synecocyst
159	35	47.9	802	1	PEFC_SALTY	P37868	salmonella	232	34	46.6	483	1	VGLY_TACV5	P13542	tacaribe vi
160	35	47.9	812	1	RIRI_DROME	P48591	drosophila	233	34	46.6	483	1	VGLY_TACVT	P13840	tacaribe vi
161	35	47.9	930	1	Y0G2 YEAST	P40367	saccharomyc	234	34	46.6	495	1	VGLY_TACCV	P18141	tacaribe vi
162	35	47.9	925	1	YES9_SCHPO	O13776	schizosacch	235	34	46.6	522	1	STA_RICCO	Q10710	ricinus com
163	35	47.9	1068	1	HMHD_ASPTT	Q9Y7D2	aspergillus	236	34	46.6	523	1	STC_RICCO	Q41144	ricinus com
164	35	47.9	1115	1	TBC2_CHLRE	O8VXP3	chlamydomon	237	34	46.6	532	1	SFER_STRPU	P16264	strongyloe
165	35	47.9	1356	1	Y195_HUMAN	Q12767	homo sapien	238	34	46.6	532	1	DCPI_SCHPO	Q09737	schizosacch
166	34.5	47.3	234	1	EMB2_CAVPO	P35709	cavia porce	239	34	46.6	574	1	CHLE_HORSE	P81508	equus cabal
167	34.5	47.3	275	1	RCEL_RHORI	P10717	rhodospiril	240	34	46.6	640	1	UL06_HSV7J	P52455	human herpe
168	34.5	47.3	367	1	YM11_PARTE	P15612	paramycium	241	34	46.6	644	1	VP4_BTIV1	P33428	bluetongue
169	34.5	47.3	420	1	Y439_PVRHC	O58758	pyrococcus	242	34	46.6	644	1	VP4_BTIV2A	P33427	bluetongue
170	34.5	47.3	443	1	CGLH_XANMA	Q37126	xanthomonas	243	34	46.6	654	1	VP4_BTIV10	P07132	bluetongue
171	34.5	47.3	446	1	LAMB_ECO57	O8X5W7	escherichia	244	34	46.6	666	1	Y435_MYCPN	P75079	mycoplasma
172	34.5	47.3	446	1	LAMB_ECOLI	Q8CV14	escherichia	245	34	46.6	712	1	S21F_MOUSE	Q9N9B5	mus musculus
173	34.5	47.3	446	1	LAMB_ECOLI	P02943	escherichia	246	34	46.6	715	1	S21F_MOUSE	Q9EPZ7	rattus norv
174	34.5	47.3	452	1	LAMB_SALTI	Q821T9	salmonella	247	34	46.6	724	1	NOSR_PSEST	Q00790	pseudomonas
175	34.5	47.3	452	1	LAMB_SALTY	P26466	salmonella	248	34	46.6	788	1	EXGL_COCCA	P49426	cochliobol
176	34.5	47.3	485	1	ENT_ENTCO	P81007	enterolobiu	249	34	46.6	854	1	ENV_SIVCZ	P17281	chimpanzee
177	34.5	47.3	812	1	FTKI_NEIMA	Q9JU31	neisseria m	250	34	46.6	964	1	EMPE_CHLTR	O84877	chlamydia t
178	34.5	47.3	812	1	FTKI_NEIMA	Q9J236	neisseria m	251	34	46.6	976	1	EMPE_CHLTR	Q9P147	chlamydia m
179	34.5	47.3	2206	1	POLG_POLIM	P03299	p genome po	252	34	46.6					

BT Drosophila and analysis of three P element insertions.";
RL Nucleic Acids Res. 20:4639-4647(1992).
RN [4]
RP REVISIONS TO 2068-2148.
RX MEDLINE=95018278; PubMed=7932764;
RA Davidson J.N., Kern C.B.;
RT "Revision in sequence of CAD aspartate transcarbamylase domain of
RL Drosophila"; 243:364-366(1994).
RL J. Mol. Biol. 243:364-366(1994).
CC -!- FUNCTION: This protein is a "fusion" protein encoding four
CC enzymatic activities of the pyrimidine pathway (GATase, CPSase,
CC ATCase and DHase).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -!- CATALYTIC ACTIVITY: Carbamoyl phosphate + L-aspartate = phosphate
CC + N-carbamoyl-L-aspartate.
CC -!- CATALYTIC ACTIVITY: (S)-dihydroorotate + H(2)O = N-carbamoyl-L-
CC aspartate.
CC -!- COFACTOR: Dihydroorotase: binds 1 zinc ion per subunit
CC (Potential).
CC -!- PATHWAY: Pyrimidine biosynthesis; first step.
CC -!- PATHWAY: Pyrimidine biosynthesis; second step.
CC -!- PATHWAY: Pyrimidine biosynthesis; third step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- MISCELLANEOUS: GATase (glutamine amidotransferase) and CPSase
CC (carbamoyl phosphate synthase) form together the glutamine-
CC dependent CPSase (GD-CPSase) (EC 6.3.5.5).
CC -!- SIMILARITY: THE CPSASE DOMAIN IS SIMILAR TO OTHER CPASES.
CC -!- SIMILARITY: In the central section; belongs to the DHase family.
CC -!- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X34813; CAA28502.1; -;
CC EMBL; X33875; CAA27509.1; -;
CC EMBL; X33876; CAA27510.1; ALT SEQ.
CC EMBL; X33877; CAA27511.1; ALT SEQ.
CC EMBL; X33878; CAA27512.1; -;
CC EMBL; X33879; CAA27513.1; -;
CC EMBL; M37783; AAA28873.1; -;
CC EMBL; S74010; AAB32204.1; -;
CC PIR; A29106; QZFF.
CC RSP; P00479; 3CSU.
CC FlyBase; FBgn0003189; r.
CC InterPro; IPR006680; Amidohydro.1.
CC InterPro; IPR006130; Asp/Orn_CoTransf.
CC InterPro; IPR02082; Asp_carbamyltransf.
CC InterPro; IPR06275; CarA_L_glu.
CC InterPro; IPR001317; CP synthGATase.
CC InterPro; IPR002474; CP synthsmall.
CC InterPro; IPR005483; CPase_L.
CC InterPro; IPR005479; CPase_L_D2.
CC InterPro; IPR005480; CPase_L_D3.
CC InterPro; IPR005481; CPase_L_N.
CC InterPro; IPR000991; GATase_I.
CC InterPro; IPR004362; MGS like.
CC InterPro; IPR006131; OTCace_O.
CC InterPro; IPR006132; OTCace_P.
CC InterPro; IPR002195; Pept_M38_nph.
CC Pfam; PF01979; Amidohydro.1; 1.
CC Pfam; PF00289; CPSase_L_chain; 2.
CC Pfam; PF02786; CPSase_L_D2; 2.
CC Pfam; PF02787; CPSase_L_D3; 1.
CC Pfam; PF00986; CPSase_sm_chain; 1.
CC Pfam; PF00117; GATase; 1.
CC Pfam; PF02142; MGS; 1.
CC Pfam; PF00185; OTCace; 1.

DR PF02729; OTCace_N; 1.
DR PRINTS; PR00100; AOTCase.
DR PRINTS; PR00098; CPSASE.
DR PRINTS; PR00099; CPSGATASE.
DR PRINTS; PR00096; GATASE.
DR TIGRFAMs; TIGR00670; asp_carb_tr; 1.
DR TIGRFAMs; TIGR01369; CPSaseII_lrg; 1.
DR PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.
DR PROSITE; PS00442; GATASE_TYPE_1; 1.
DR PROSITE; PS00482; DIHYDROOROTASE_1; FALSE_NEG.
DR PROSITE; PS00483; DIHYDROOROTASE_2; 1.
DR PROSITE; PS00866; CPSASE_1; 1.
DR PROSITE; PS00867; CPSASE_2; 1.
KW Pyrimidine biosynthesis; Ligase; Transferase; Hydrolase; Zinc;
KM Multifunctional enzyme.
FT DOMAIN 1 381 GATASE (GLUTAMINE AMIDOTRANSFERASE).
FT DOMAIN 382 420 LINKER.
FT DOMAIN 421 1473 CPSASE (CARBAMOYL-PHOSPHATE SYNTHASE).
FT DOMAIN 1474 1487 LINKER.
FT DOMAIN 1488 1802 DHOASE (DIHYDROOROTASE).
FT DOMAIN 1803 1917 LINKER.
FT DOMAIN 1918 2236 ATCase (ASPARTATE TRANS-CARBAMYLASE).
FT ACT_SITE 274 274 GATASE (BY SIMILARITY).
FT METAL 1489 1489 ZINC (POTENTIAL).
FT METAL 1491 1491 ZINC (POTENTIAL).
FT CONFLICT 2075 2104 WRLEKTRDALAGPPADPQCEPAVMWRD (IN REF.
FT 1).
FT CONFLICT 2131 2145 NVLPDVLVMTRIQ -> ECAARHGCAALHDSHS (IN
FT REF. 1).
FT SQ SEQUENCE 2236 AA; 249238 MW; 85A3D7CBA82E36A1 CRC64;
Query Match 59.6%; Score 43.5; DB 1; Length 2236;
Best Local Similarity 41.2%; Pred. No. 1.7e+02;
Matches 7; Conservative 1; Mismatches 2; Indels 7; Gaps 1;
QY 1 EWTN-----IWNWAK 10
||| ||||
DB 88 EWTEGRRPGQPPWWAR 104
RESULT 3
FDFT USTMA STANDARD; PRT; 572 AA.
AC Q92459;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Farnesyl-diphosphate farnesyltransferase (EC 2.5.1.2.) (Squalene
DE synthetase) (SQS) (SS) (FPP:FPP farnesyltransferase).
GN ERG9.
OS Ustilago maydis (Smut fungus).
OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC Ustilaginomycetidae; Ustilaginales; Ustilagaceae; Ustilago.
OX NCBI_TaxID=5270;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IMI 103761;
RA Corran A.J.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: 2 farnesyl diphosphate = diphosphate +
CC presqualene diphosphate.
CC -!- CATALYTIC ACTIVITY: Presqualene diphosphate + NADPH = diphosphate
CC + squalene + NADP(+).
CC -!- COFACTOR: Magnesium.
CC -!- PATHWAY: Critical branch point enzyme of isoprenoid and
CC cholesterol biosynthesis.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
CC reticulum (By similarity).
CC -!- SIMILARITY: Belongs to the phytoene/squalene synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration


```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X99718; CAA66054.1; -
DR InterPro; IPR002060; Squ/phyt_synthese.
DR InterPro; IPR006449; Squal synth.
DR InterPro; IPR008949; Terpenoid synth.
DR Pfam; PF00494; SQS_PSY; 1
DR TIGRFAMs; TIGR01559; squa1_synth; 1
DR PROSITE; PS01044; SQUALEN PHYTOEN SYN 1; 1.
DR PROSITE; PS01045; SQUALEN PHYTOEN SYN 2; 1.
KW Multifunctional enzyme; Transferase; Oxidoreductase; NADP; Magnesium;
KW Isoprene biosynthesis; Cholesterol biosynthesis; Transmembrane;
KW Endoplasmic reticulum
FT TRANSMEM 316 336 POTENTIAL.
FT TRANSMEM 492 512 POTENTIAL.
SQ SEQUENCE 572 AA; 65508 MW; A38A40E29D885C0D CRC64;

Query Match 58.9%; Score 43; DB 1; Length 572;
Best Local Similarity 57.1%; Pred. No. 52;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 WTWIWW 8
DB 512 WEIWW 518

RESULT 4
YQAQ SCHPO STANDARD; PRT; 688 AA.
AC Q10109;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C1866.10 in chromosome 1.
GN SPAC1866.10.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OC NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeil C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Welter J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer B., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reichardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Paga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey P., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
```

```
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z68198; CAA92388.1; -
DR PIR; T37923; T37923.186.10; -
DR GeneDB; SPombe; SPAC1866.10; -
KW Hypothetical protein; Transmembrane.
KW TRANSMEM 318 338 POTENTIAL.
FT TRANSMEM 547 567 POTENTIAL.
SQ SEQUENCE 688 AA; 78178 MW; 3CB29ED7B7A5DEA2 CRC64;

Query Match 58.9%; Score 43; DB 1; Length 688;
Best Local Similarity 62.5%; Pred. No. 62;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNIWW 8
DB 658 EWTNIWW 665

RESULT 5
CDGL HUMAN STANDARD; PRT; 938 AA.
AC Q9Y5F7; Q9Y5C3;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Protocadherin gamma C4 precursor (PCDH-gamma-C4).
GN PCDHGC4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TSSOE=Brain;
RX MEDLINE=99308636; PubMed=10380929;
RA Wu Q., Maniatis T.;
RT "A striking organization of a large family of human neural cadherin-
RL like cell adhesion genes.";
RL Cell 97:779-790(1999).
CC -!- FUNCTION: Potential calcium-dependent cell-adhesion protein. May
CC be involved in the establishment and maintenance of specific
CC neuronal connections in the brain.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9Y5F7-1; Sequence=Displayed;
CC Name=2; Synonyms=Short;
CC IsoId=Q9Y5F7-2; Sequence=VSP_008700, VSP_008701;
CC -!- SIMILARITY: Contains 6 cadherin domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF152338; AAD43732.1; -
DR EMBL; AF152525; AAD43785.1; -
DR Genbank; HGNC:8717; PCDHGC4.
DR MIM; 606305; -
```

```
DR MIM: 604968; --
DR InterPro: IPR002126; Cadherin.
DR Pfam: PF00328; cadherin; 5.
DR PRINTS: PRC0205; CADHERIN.
DR SMART: SM00112; CA; 6.
DR PROSITE: PS00232; CADHERIN 1; 5.
DR PROSITE: PS0268; CADHERIN 2; 6.
DR Calcium-binding: Cell adhesion; Glycoprotein; Signal; Repeat;
KW Transmembrane; Multigene family; Alternative splicing.
KW SIGNAL 1 29
FT CHAIN 30 938 PROTOCADHERIN GAMMA C4.
FT DOMAIN 30 692 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 693 713 POTENTIAL.
FT DOMAIN 714 938 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 1 133 CADHERIN 1.
FT DOMAIN 2 242 CADHERIN 2.
FT DOMAIN 3 350 CADHERIN 3.
FT DOMAIN 4 455 CADHERIN 4.
FT DOMAIN 5 565 CADHERIN 5.
FT DOMAIN 6 676 CADHERIN 6.
FT CARBOHYD 265 265 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 815 871 QAPNTDWRFSQRPFGTSGSDGDTGTPNNQFOTEMDQ
AMTASAEZADGSSST -> VRPLQAQQLGEGYSDPGIMLG
HVLSTGLSVSAHSVDVTFVRGNVYVDAVLCNCFVN (in
isoform 2).
FTId=VSP_008700.
/FTId=VSP_008701.
FT VARSPLIC 872 938 Missing (in isoform 2).
FT SEQUENCE 938 AA; 101213 MW; 99820B82A2F1BECC CRC64;
Query Match 58.3%; Score 43; DB 1; Length 938;
Best Local Similarity 75.3%; Pred. No. 84;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 WTNIWPA 9
Db 8 WTEIWEWA 15
RESULT 6
ID YID9 YEAST STANDARD; PRT; 473 AA.
AC P40533;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical 54.9 kDa protein in CHR5-NOT3 intergenic region.
GN YI039W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9288C / AB972;
RX MEDLINE=9733266; PubMed=9169870;
RA Churcher C.M., Bowman S., Badcock K., Bankier A., Brown D.,
RA Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,
RA Harris D.E., Horsnell T., Hunt S., Jagels K., Jones M., Lye G.,
RA Kye S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N.,
RA Skelton J., Smith V., Walsh S., Whitehead S., Barrell B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX.";
RL Nature 387:84-87(1997).
CC -!- SIMILARITY: SOME, TO S.POMSE SPAC23A1.02C.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z46861; CA86912.1; --
CC FIR: S49939; S49939.
CC Germline: I39576; --
CC SCD: S0001301; YI039W.
CC InterPro: IPR004843; M-peptidase.
CC Pfam: PF00149; Metallophos; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 10 30 POTENTIAL.
FT TRANSMEM 453 473 POTENTIAL.
SQ SEQUENCE 473 AA; 54927 MW; 3D8AF321C9BAFDE8 CRC64;
Query Match 58.2%; Score 42.5; DB 1; Length 473;
Best Local Similarity 36.3%; Pred. No. 50;
Matches 7; Conservative 3; Mismatches 0; Indels 9; Gaps 1;
QY 1 EWT-----NIWMAK 10
Db 439 EWTFLCPFAIQHWWPAK 457
RESULT 7
ID YD71 SCHPO STANDARD; PRT; 720 AA.
AC Q10326;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 23-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C32A11.01 in chromosome I.
GN SPAC32A11.01.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Monte S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Gymnopoulos B.,
RA Weigtens I., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl E., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Drenth E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochard M., Gaillardin C., Tallada V.A., Garzon A., Rhode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z69736; CAA93698.1; -.
DR PIR; T38647; T38647.
DR GenBank; SPombe; SPAC32A11.01; -.
KW Hypothetical protein.
SQ SEQUENCE 720 AA; 82456 MW; D4DD9E6FC6B5604C CRC64;

Query Match
Best Local Similarity 58.2%; DB 1; Length 720;
Matches 7; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 3 TINIWM-WAK 10
DB 398 TINIWMWAE 406
|||||

RESULT 8
DNJ2 CAEEL STANDARD; PRT; 337 AA.
AC Q17433;
DT 28-FEB-2003 (Rel. 41, Last Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dnaj homolog dnj-2 precursor.
GN DNJ-2 OR B0335.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA White S.;
CC -1- SIMILARITY: Contains 1 J domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z73102; CAA97409.1; -.
DR PIR; T18654; T18654.
DR WormRep; B0035.2; CE05161.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ; 1.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS50076; DnaJ_2; 1.
KW Signal; Chaperone.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 337 DnaJ HOMOLOG DNJ-2.
FT DOMAIN 36 105 J DOMAIN.
SQ SEQUENCE 337 AA; 39992 MW; 03CB56511FE0185D CRC64;

Query Match
Best Local Similarity 56.2%; DB 1; Length 337;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNIWMW 8
DB 238 KWTALWY 245
|||||

RESULT 9
SUMI_MOUSE STANDARD; PRT; 372 AA.
ID SUMI_MOUSE
AC QBR0F3;
DT 10-OCT-2003 (Rel. 42, Created)

or send an email to license@isb-sib.ch).
10-OCT-2003 (Rel. 42, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
Sulfatase modifying factor 1 precursor (C-alpha-formylglycine-
generating enzyme 1).
GN SUMF1 OR FGE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Retina;
MEDLINE=22388257; Pubmed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Joquellano N.A., Peters G.J., Abramson R.D., Mullaney S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Gramwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- FUNCTION: Converts newly synthesized inactive sulfatases to their
active form by modifying an active site cysteine residue to 2-
amino-3-oxopropionic acid. Known substrates include GALNS, ARSA,
STS and ARSE (By similarity).
CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumenal protein (By
similarity).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; BC026981; AAH26981.1; ALT_INIT.
DR MGD; MGI:1889844; AI463102.
DR InterPro; IPR005532; DUF323.
DR Pfam; PF03781; DUF323; 1.
DR Signal; Endoplasmic reticulum; Glycoprotein.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 372 SULFATASE MODIFYING FACTOR 1.
FT CARBOHYD 139 139 N-LINKED (GLCNAC ) (POTENTIAL).
SQ SEQUENCE 372 AA; 40659 MW; 3C96D48B4291068 CRC64;

Query Match
Best Local Similarity 56.2%; DB 1; Length 372;
Best Local Similarity 71.4%; Pred. No. 63;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNIWMW 7
DB 298 EWTSDMW 304
|||||

RESULT 10
SUMI_HUMAN STANDARD; PRT; 374 AA.
ID SUMI_HUMAN
AC Q8NBK3; Q96AK5; Q96DK8;
DT 10-OCT-2003 (Rel. 42, Created)
```

DT 10-OCT-2003 (rel. 42, Last sequence update)
DT 10-OCT-2003 (rel. 42, Last annotation update)
DE Sulfatase modifying factor 1 precursor (C-alpha-formylglycine-
DE generating enzyme 1).
GN SUMF1 OR FGE.
CS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RX (1)
RP SEQUENCE FROM N.A. (ISOFORM 1), VARIANTS MSD VAL-279; ARG-336;
RP GLN-349 AND TRP-349; AND VARIANT ASN-63.
RX MEDLINE=22642673; PubMed=12757705;
RA Dierks T., Schmidt B., Borissenko L.V., Peng J., Preusser A.,
RA Mariappan N., von Figura K.,
RT Multiple sulfatase deficiency is caused by mutations in the gene
RT encoding the Homo sapiens C-alpha-formylglycine-generating enzyme.";
RL Cell 113:435-444(2003).
RN (2)
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Gastric mucosa;
RA Kinomiyu K., Nagatsuna M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikawa E., Omura Y., Abe K., Kamihara K.,
RA Nakagawa K., Sato K., Tanikawa M., Yamazaki M., Suzuki Y., Hata H.,
RA Irie R., Otsuki T., Sato H., Mshikawa T., Sugiyama A., Kawakami B.,
RA Nagai K., Isogai T., Sugano S.;
RT "NEO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Placenta;
RA Ota T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayashi K., Ishii S.,
RA Saito K., Yamamoto J., Wakamatsu A., Nagai T., Nakamura Y.,
RA Nagahari K., Sugano S., Isogai T.;
RT "HRI human cDNA sequencing project.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN (4)
RP SEQUENCE OF 234-374 FROM N.A.
RC TISSUE=Prostate;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L.H., Shenren C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Diatchenko L., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN (5)
RP VARIANTS MSD PRO-155; TYR-218; ARG-336; CYS-345; PRO-348; GLN-349 AND
RP TRP-349 AND FUNCTION
RX MEDLINE=22642674; PubMed=12757706;
RA Cosma M.P., Pepe S., Annunziata I., Newbold R.F., Grompe M.,
RA Parenti G., Ballabio A.;
RT "The multiple sulfatase deficiency gene encodes an essential and
RT limiting factor for the activity of sulfatases.";
RL Cell 113:445-456(2003).
CC -!- FUNCTION: Converts newly synthesized inactive sulfatases to their
CC active form by modifying an active site cysteine residue to 2-
CC amino-3-oxopropionic acid. Known substrates include GALNS, ARSA,

CC STS and ARSE.
CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum luminal protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q8NBK3-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8NBK3-2; Sequence=VSP_007877;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Ubiquitous. Highly expressed in kidney, lung,
CC pancreas and liver. Detected at lower levels in leukocytes, lung,
CC placenta, small intestine, skeletal muscle and heart.
CC -!- PTM: N-glycosylated.
CC -!- DISEASE: Defects in SUMF1 are the cause of multiple sulfatase
CC deficiency (MSD) [MIM:272200]. MSD is a clinically and
CC biochemically heterogeneous disorder caused by the simultaneous
CC impairment of all sulfatases, due to defective post-translational
CC modification and activation. It combines features of individual
CC sulfatase deficiencies such as metachromatic leukodystrophy,
CC mucopolysaccharidosis, chondrodysplasia punctata, hydrocephalus,
CC ichthyosis, neurologic deterioration and developmental delay.
CC Inheritance is autosomal recessive.
CC -!- SIMILARITY: Belongs to the sulfatase modifying factor family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AY208752; BAB34683.1; -
CC EMBL; AK057983; BAB71625.1; -
CC EMBL; AK075459; BAC11634.1; -
CC EMBL; BC017005; AAH17005.1; -
CC MIM; 607939; -
CC MIM; 272200; -
CC InterPro; IPR005532; DUF323.
CC Pfam; PF03781; DUF323; 1.
KW Signal; Endoplasmic reticulum; Glycoprotein; Polymorphism;
KW Disease mutation; Mucopolysaccharidosis; Metachromatic leukodystrophy;
KW Alternative splicing.
FT SIGNAL 1 33 POTENTIAL SULFATASE MODIFYING FACTOR 1.
FT CHAIN 34 374 N-LINKED GLUCNA... (POTENTIAL).
FT CARBOHYD 141 141 Missing (in isoform 2).
FT VARSPPLIC 1 90 /FTID=VSP_007877.
FT VARIANT 63 63 S -> N.
FT VARIANT 155 155 /FTID=VAR_016052.
FT VARIANT 155 155 S -> P (in MSD).
FT VARIANT 218 218 /FTID=VAR_016053.
FT VARIANT 218 218 C -> Y (in MSD).
FT VARIANT 279 279 A -> V (in MSD).
FT VARIANT 279 279 /FTID=VAR_016054.
FT VARIANT 336 336 C -> R (in MSD).
FT VARIANT 336 336 /FTID=VAR_016055.
FT VARIANT 345 345 R -> C (in MSD).
FT VARIANT 345 345 /FTID=VAR_016056.
FT VARIANT 348 348 A -> P (in MSD).
FT VARIANT 348 348 /FTID=VAR_016057.
FT VARIANT 349 349 R -> Q (in MSD).
FT VARIANT 349 349 /FTID=VAR_016058.
FT VARIANT 349 349 R -> W (in MSD).
FT VARIANT 349 349 /FTID=VAR_016059.
FT CONFLICT 124 124 L -> F (IN REF. 1 AND 2).
FT CONFLICT 264 264 E -> D (IN REF. 3).
SQ SEQUENCE 374 AA; 40522 MW; B64F2F5204C8CA43 CRC64;
Query Match 56.2%; Score 41; DB 1; Length 374;
Best Local Similarity 71.4%; Pred. No. 63;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 EWTNIWW 7
Db 300 EWTSDWW 306

RESULT 11
YLD2_CAEEL STANDARD; PRT; 493 AA.
AC Q03567;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein C38C10.2 in chromosome III.
GN C38C10.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;
OC Rhabditiidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol NZ;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Connell M., Copey T., Cooper J., Coulson A.,
RA Bonfield J., Burton J., Durbin R., Favello A., Fraser A.,
RA Craxton M., Dear S., Du Z., Durbin R., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kersey P., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Cooper A., Saunders D., Showkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonnenhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaughan M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wooldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP REVISIONS.
RA Jones S.J.M.;
RA Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: Belongs to the sodium/phosphate cotransporter family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z19153; CAA79549.1; -.
CC PIR; G88553; G88553.
CC WormPep; C38C10.2; CE03647.
CC InterPro; IPR007114; MFS.
CC InterPro; IPR005828; Sub_transporter.
CC Pfam; PF00083; sugat_tr; 1.
CC PROSITE; PS00850; MFS; 1.
CC KW Hypothetical protein; Transport; Symport; Sodium transport;
CC Transmembrane; Glycoprotein.
FT TRANSMEM 10 30 POTENTIAL.
FT TRANSMEM 85 105 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 144 164 POTENTIAL.
FT TRANSMEM 175 195 POTENTIAL.
FT TRANSMEM 205 225 POTENTIAL.
FT TRANSMEM 272 292 POTENTIAL.
FT TRANSMEM 311 331 POTENTIAL.
FT TRANSMEM 348 368 POTENTIAL.
FT TRANSMEM 375 395 POTENTIAL.
FT TRANSMEM 406 426 POTENTIAL.
FT TRANSMEM 441 461 POTENTIAL.

SQ SEQUENCE 493 AA; 52737 MW; 60792715D32553DD CRC64;
Query Match 56.2%; Score 41; DB 1; Length 493;
Best Local Similarity 57.1%; Pred. No. 83;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WTNINWW 8
Db 218 WTAUVWY 224

RESULT 12
SYK_AERPE STANDARD; PRT; 562 AA.
AC Q9YFT9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Lysyl-tRNA synthetase (EC 6.1.1.6) (Lysine--tRNA ligase) [LYSRS].
GN LYSS OR APE0161.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hcosoyana A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
CC -!- CATALYTIC ACTIVITY: ATP + L-lysine + tRNA(Lys) = AMP + diphosphate
CC + L-lysyl-tRNA(Lys).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP000058; BAA79072.1; -.
CC PIR; F72771; F72771.
CC HAMAP; MF_00177; -.
CC InterPro; IPR002904; Lys tRNA-synt lc.
CC InterPro; IPR008925; tRNA-synt bind.
CC InterPro; IPR001412; tRNA-synt 1.
CC Pfam; PF01921; tRNA-synt 1f; 1.
CC TIGRFAMs; TIGR00467; lyss_arch; 1.
CC PROSITE; PS00178; AA tRNA_LIGASE_1; FALSE NEG.
CC KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Complete proteome.
FT SITE 50 58 "HIGH" REGION.
FT SITE 305 309 "KMSKS" REGION.
SQ SEQUENCE 562 AA; 65114 MW; 753664E2937FBF27 CRC64;
Query Match 56.2%; Score 41; DB 1; Length 562;
Best Local Similarity 57.1%; Pred. No. 94;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNIWW 7
Db 246 EWAGVWV 252
```

```

RESULT 13
RX 367 HABIN STANDARD; PRT; 303 AA.
AC Q57065; C05019;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein H13367.
GN H13367.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fieischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uitterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.B., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
RL Science 269:496-512(1995).
CC -!- SIMILARITY: WEAK, TO E.COLI YFGA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; U32721; AAC22025.1; -.
DR PIR; A64150; A64150.
DR TIGR; H30367; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 303 AA; 34011 MW; 4A3FEFB363D5653C CRC64;
Query Match 54.8%; Score 40; DB 1; Length 303;
Best Local Similarity 66.7%; Pred. No. 70;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 3 TNLWVW 8
Db 124 TGLWVW 129
|:|:|
|:|:|
RESULT 14
POLG CXA24 STANDARD; PRT; 2214 AA.
AC P36230;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-OCT-2003 (Rel. 42, Last annotation update)
DE Genome polyprotein [Contents: Coat protein VP4 (P1A); Coat protein VP2
DE (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Core genome-
DE P2A; Core protein P2B; Core protein P2C; Core protein P3A; Genome-
DE linked protein VP6 (P3B); Picornain 3C (EC 3.4.22.28) (Protease 3C)
DE (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].
OS Coxsackievirus A24 (strain EH24/70).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Enterovirus.
OX NCBI_TaxID=36404;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=92271460; PubMed=1317075;
RA Supanaranond K., Takeda N., Yamazaki S.;
RT "The complete nucleotide sequence of a variant of Coxsackievirus A24,
RT an agent causing acute hemorrhagic conjunctivitis.";
RL Virus Genes 6:149-158(1992).
CC -!- FUNCTION: It is thought that the P2C protein attaches to vesicular
CC membranes and is associated with viral RNA synthesis.
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC O/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -!- CATALYTIC ACTIVITY: Selective cleavage of Gln-I-Gly bond in the
CC poliovirus polyprotein. In other picornavirus reactions Glu may be
CC substituted for Gln, and Ser or Thr for Gly.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,
CC each of which is composed of one copy each of proteins VP1, VP2,
CC VP3, and VP4.
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; D90457; -. NOT_ANNOTATED_CDS.
DR PIR; A48548; A48548.
DR HSSP; P03299; 1AR7.
DR MEROPS; C03.001; -.
DR MEROPS; C03.020; -.
DR InterPro; IPR004004; Calici_pol_hel.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR000199; Pept_3C_picorn.
DR InterPro; IPR000081; Peptidase_C3.
DR InterPro; IPR003138; Pico_P1A.
DR InterPro; IPR002527; Pico_P2B.
DR InterPro; IPR001676; Rbv_P2B.
DR InterPro; IPR000605; RNA_helicase.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR001205; RNA_pol_P3D.
DR InterPro; IPR007094; RNA_pol_PSwir.
DR InterPro; IPR008975; Viral_cap_coat.
DR Pfam; PF00548; Cys-protease-3C; 1.
DR Pfam; PF02226; Pico_P1A; 1.
DR Pfam; PF00947; Pico_P2A; 1.
DR Pfam; PF01552; Pico_P2B; 1.
DR Pfam; PF00073; rhv; 3.
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR Pfam; PF00910; RNA_helicase; 1.
DR PRINTS; PR00918; CALICIVIRUSNS.
DR ProDom; PD001125; Cys_protease_3C; 1.
DR ProDom; PD001306; Pico_P2A; 1.
DR ProDom; PD001274; Pico_P2B; 1.
DR ProDom; PD001274; Pico_P2B; 1.
KW Polyprotein; Coat protein; Core protein; Transferase;
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate;
KW Lipoprotein.
FT CHAIN 2 69 COAT PROTEIN VP4.
FT CHAIN 70 340 COAT PROTEIN VP2.
FT CHAIN 341 577 COAT PROTEIN VP3.
FT CHAIN 578 888 COAT PROTEIN VP1.
FT CHAIN 889 1035 CORE PROTEIN P2A.
FT CHAIN 1036 1132 CORE PROTEIN P2B.
FT CHAIN 1133 1461 CORE PROTEIN P2C.
FT CHAIN 1462 1548 CORE PROTEIN P3A.
FT CHAIN 1549 1570 GENOME-LINKED PROTEIN VP6.
FT CHAIN 1571 1753 PICORNAIN 3C.
FT CHAIN 1754 2214 RNA-DIRECTED RNA POLYMERASE.
FT CHAIN 2 2 N-myristoyl glycine (in host) (By
FT LIPID 1717 1717 PROTEASE (POTENTIAL)).
FT ACT_SITE 1717 1717

```

```
FT ACT_SITE 1731 1731 PROTEASE (POTENTIAL).
SQ SEQUENCE 2214 AA; 247212 MW; E0DD74569E1B22B8 CRC64;

Query Match 54.1%; Score 39.5; DB 1; Length 2214;
Best Local Similarity 45.5%; Pred. No. 5,6e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 3; Gaps 1;

Qy 1 EWTNI---WWW 8
   :||:
Db 139 QWTSTSGWW 149

RESULT 15
VNSA WHVW6 STANDARD; PRT; 282 AA.
AC P11233;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Major surface antigen precursor.
GN S.
OS Woodchuck hepatitis virus w64 (isolate pWS23).
OC Viruses; Retroid viruses; Hepadnaviridae; Orthohepadnavirus.
OX NCBI_TaxID=10436;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=87219879; PubMed=3582379;
RX Ettemble J., Moeroy T., Trepo C., Tiollais P., Buedia M.-A.;
RT "Nucleotide sequence of the woodchuck hepatitis virus surface antigen
RL mRNAs and the variability of three overlapping viral genes.";
RL Gene 50:207-214 (1986).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M15954; AAA69574.1; -.
DR PIR; B29498; SAVL64.
DR InterPro; IPR000349; Hepvir_surfag.
DR Pfam; PFC0695; VNSA; 1.
DR AcciGen.
RW PROPEP 1 60
FT CHAIN 61 282 MAJOR SURFACE ANTIGEN.
SQ SEQUENCE 282 AA; 32053 MW; 7500A89BE4097277 CRC64;

Query Match 53.4%; Score 39; DB 1; Length 282;
Best Local Similarity 40.0%; Pred. No. 88;
Matches 6; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

Qy 2 WTNI-----WWW 8
   :||:
Db 80 WTKIPTTAQNLDWW 94

RESULT 16
RFBB RHISN STANDARD; PRT; 350 AA.
AC P55462;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Probable dTDP-glucose 4,6-dehydratase (EC 4.2.1.46).
GN Y4GF.
OS Rhizobium sp. (strain NGR234).
CG Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401 (1997).
CC -!- CATALYTIC ACTIVITY: dTDP-glucose = dTDP-4-dehydro-6-deoxy-D-
CC glucose + H(2)O.
CC -!- COFACTOR: NAD.
CC -!- PATHWAY: dTDP-L-rhamnose biosynthesis.
CC -!- SIMILARITY: BELONGS TO THE SUGAR EPIMERASE FAMILY. DTDG-GLUCOSE
CC DEHYDRATASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000074; AAB91680.1; -.
DR HSP; P27830; 1BXK.
DR InterPro; IPR005888; dTDP_gluc_dehyd.
DR InterPro; IPR001509; Epimerase_Dh.
DR Pfam; PFO1370; Epimerase; 1.
DR TIGRfams; TIGR01181; dTDP-gluc_dehyd; 1.
KW Hypothetical protein; Lyase; NAD; Plasmid.
FT NP_BIND 7 13
SQ SEQUENCE 350 AA; 39664 MW; B6EF2B437DD751A4 CRC64;

Query Match 53.4%; Score 39; DB 1; Length 350;
Best Local Similarity 60.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

Qy 1 EW--TNIWWW 8
   :||:
Db 322 EWLENGWW 331

RESULT 17
GUN3 HUMIN STANDARD; PRT; 388 AA.
ID GUN3 HUMIN STANDARD; PRT; 388 AA.
AC Q12624; Q12620;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endoglucanase 3 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase 3)
DE (Cellulase 3).
GN CMG3.
OS Humicola insolens.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
OX NCBI_TaxID=34413;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94247364; PubMed=8190078;
RA Dalboege H., Hansen H.P.H.;
RT "A novel method for efficient expression cloning of fungal enzyme
RT genes.";
RL Mol. Gen. Genet. 243:253-260 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN-IFO 9854;
RC Takashima S., Nakamura A., Masaki H., Uozumi T.;
RT "Cloning, sequencing and expression of the thermostable cellulase
RT gene of Humicola grisea.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- BIOTECHNOLOGY: Used as a detergent cellulase. Sold under the name
CC Celluzyme by Novozymes. This special cellulase has three effects:
CC colour brightening, softening and removal of particulate soil. The
CC overall effect is that it helps to preserve the nice appearance of
```



```
FT PROPEP      1      204
FT CHAIN       205    426      MAJOR SURFACE ANTIGEN.
FT CARBOHYD    32      32      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD    147    147      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD    346    346      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE    426 AA; 48403 MW; B90988B556468097 CRC64;

Query Match
Best Local Similarity 53.4%; Score 39; DB 1; Length 426;
Matches 6; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

QY 2 WTNI-----WWW 8
DB 224 WTKILTIAQSLDWW 238

RESULT 22
VMSA_HPBGS
ID VMSA_HPBGS STANDARD; PRT; 428 AA.
AC P03144;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-AUG-1990 (Rel. 15, Last annotation update)
DE Major surface antigen precursor.
GN S.
OS Ground squirrel hepatitis virus (GSV).
OC Viruses; Retroviral viruses; Hepadnaviridae; Orthohepadnavirus.
ON NCBI_TaxID=10406;
RN [1]
RP SEQUENCE FROM N.A.
RA Seeger C., Ganem D., Varmus H.E.;
RX MEDLINE=84267998; PubMed=6086950;
RT "Nucleotide sequence of an infectious molecularly cloned genome of
ground squirrel hepatitis virus."
RL J. Virol. 51:367-375 (1984).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; K02715; AAA46757.1; ALT_INIT.
DR PIR; A03703; SAVLS.
DR InterPro; IPR000349; Hepvir_surfa.
DR Pfam; PF00695; VMSA; 1.
KW Antigen.
FT PROPEP      1      206      MAJOR SURFACE ANTIGEN.
FT CHAIN       207    428      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD    149    149      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD    348    348      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE    428 AA; 48382 MW; 5E84114BC33862D7 CRC64;

Query Match
Best Local Similarity 53.4%; Score 39; DB 1; Length 428;
Matches 6; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

QY 2 WTNI-----WWW 8
DB 226 WTKILTIAQSLDWW 240

RESULT 21
VMSA_HPBGS
ID VMSA_HPBGS STANDARD; PRT; 431 AA.
AC P12910;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Major surface antigen precursor.
GN S.
OS Woodchuck hepatitis virus 7 (WHV 7).
OC Viruses; Retroviral viruses; Hepadnaviridae; Orthohepadnavirus.
ON NCBI_TaxID=10432;
RN [1]
RP SEQUENCE FROM N.A.
RA Cohen J.I., Miller R.H., Rosenblum B., Denniston K., Gerin J.L.,
Purcell R.H.;
RX MEDLINE=86103359; PubMed=3336938;
RT "Sequence comparison of woodchuck hepatitis virus replicative forms
shows conservation of the genome."
RL Virology 162:12-20 (1988).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; M18752; AAA46766.1; -.
DR PIR; D29969; SAVL7.
```

```
OS Woodchuck hepatitis virus 59 (WHV 59).
OC Viruses; Retroviral viruses; Hepadnaviridae; Orthohepadnavirus.
ON NCBI_TaxID=10431;
RN [1]
RP SEQUENCE FROM N.A.
RA Cohen J.I., Miller R.H., Rosenblum B., Denniston K., Gerin J.L.,
Purcell R.H.;
RX MEDLINE=86101359; PubMed=3336938;
RT "Sequence comparison of woodchuck hepatitis virus replicative forms
shows conservation of the genome."
RL Virology 162:12-20 (1988).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; M19183; AAA46762.1; -.
DR PIR; H29969; SAVL59.
DR InterPro; IPR000349; Hepvir_surfa.
DR Pfam; PF00695; VMSA; 1.
KW Antigen.
FT PROPEP      1      209      MAJOR SURFACE ANTIGEN.
FT CHAIN       210    431      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD    32      32      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD    94      94      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD    152    152      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD    351    351      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE    431 AA; 49018 MW; 5D117A1129CB0362 CRC64;

Query Match
Best Local Similarity 53.4%; Score 39; DB 1; Length 431;
Matches 6; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

QY 2 WTNI-----WWW 8
DB 229 WTKILTIAQSLDWW 243

RESULT 22
VMSA_WHV7
ID VMSA_WHV7 STANDARD; PRT; 431 AA.
AC P12909;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Major surface antigen precursor.
GN S.
OS Woodchuck hepatitis virus 7 (WHV 7).
OC Viruses; Retroviral viruses; Hepadnaviridae; Orthohepadnavirus.
ON NCBI_TaxID=10432;
RN [1]
RP SEQUENCE FROM N.A.
RA Cohen J.I., Miller R.H., Rosenblum B., Denniston K., Gerin J.L.,
Purcell R.H.;
RX MEDLINE=86103359; PubMed=3336938;
RT "Sequence comparison of woodchuck hepatitis virus replicative forms
shows conservation of the genome."
RL Virology 162:12-20 (1988).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; M18752; AAA46766.1; -.
DR PIR; D29969; SAVL7.
```

```

DR InterPro: IPR000349; Hepvir_surfag.
DR Pfam: PF00695; VMSA; 1.
KW Antigen.
FT PROPEP 1 209
FT CHAIN 210 431
FT CARBOHYD 32 32
FT CARBOHYD 94 94
FT CARBOHYD 152 152
FT CARBOHYD 351 351
SQ SEQUENCE 431 AA; 48948 MW; 50B526017EFD459 CRC64;

Query Match 53.4%; Score 39; DB 1; Length 431;
Best Local Similarity 40.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

Qy 2 WTNI-----WWW 8
Db 229 WTKILTIAQNLDWW 243

RESULT 23
VMSA WHV81
ID VMSA WHV81 STANDARD; PRT; 431 AA.
AC P17400;
DT 01-AUG-1990 (Rel. 15, Created)
DT 21-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Major surface antigen precursor.
GN S.
OS Woodchuck hepatitis virus 8 (infectious clone) (WHV 8).
OC Viruses; Retroviral viruses; Hepadnaviridae; Orthohepadnavirus.
OX NCBI_TaxID=10434;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89184524; PubMed=2928306;
RA Glorion R., Cote P.J., Hornbuckle W.B., Tennant B.C., Gerin J.L.,
RA Purcell R.H., Miller R.H.;
RT "Complete nucleotide sequence of a molecular clone of woodchuck
RT hepatitis virus that is infectious in the natural host.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:1846-1849(1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J04514; AAA46770.1; ALT_INIT.
DR PIR: B32397; SAVLW8.
DR InterPro: IPR000349; Hepvir_surfag.
DR Pfam: PF00695; VMSA; 1.
KW Antigen.
FT PROPEP 1 209
FT CHAIN 210 431
FT CARBOHYD 32 32
FT CARBOHYD 94 94
FT CARBOHYD 152 152
FT CARBOHYD 351 351
SQ SEQUENCE 431 AA; 49006 MW; F742998BCB7B95E CRC64;

Query Match 53.4%; Score 39; DB 1; Length 431;
Best Local Similarity 40.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

Qy 2 WTNI-----WWW 8
Db 229 WTKILTIAQNLDWW 243

RESULT 24
FTK6_MOUSE
ID PTK6_MOUSE STANDARD; PRT; 451 AA.
AC Q64434;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Tyrosine-protein kinase 6 (EC 2.7.1.112) (SRC-related intestinal
DE kinase).
GN PTK6 OR STK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10390;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ICR, and BALB/c; TISSUE=Intestinal crypts;
RX MEDLINE=95140424; PubMed=7838533;
RA Vasoukhin V., Serfas M.S., Siyanova E.Y., Polonskaia M.,
RA Costigan V.J., Liu B., Thomson A., Tyner A.L.;
RT "A novel intracellular epithelial cell tyrosine kinase is expressed in
RT the skin and gastrointestinal tract.";
RL Oncogene 10:349-357(1995).
RN [2]
RP SEQUENCE OF 1-77 FROM N.A.
RX STRAIN=BALB/c;
RA Siyanova E.Y.;
RT "Promoter region of the mouse sik gene.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May function as an intracellular signal transducer in
CC epithelial tissues.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- TISSUE SPECIFICITY: Expressed only in epithelial tissues,
CC including the skin and lining of the alimentary canal. Restricted
CC to the cell layers immediately above the proliferative cell zone
CC in these epithelia.
CC -!- DEVELOPMENTAL STAGE: First detected at day 15.5 of gestation in
CC the embryo, where it is expressed in the newly forming granular
CC layer of the skin. Is found in stomach at day 17.5.
CC -!- PTM: Autophosphorylated (By similarity).
CC -!- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES.
CC -!- SIMILARITY: Contains 1 SH2 domain.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U16805; AAA67929.1; -.
DR EMBL: AF016545; AAB94550.1; -.
DR HSSP: P11362; IFGK.
DR MGI: MGI:199683; Ptk6.
DR GO: GO:0005634; C:nucleus; IDA.
DR InterPro: IPR000719; Prot_kinase.
DR InterPro: IPR000980; SH2.
DR InterPro: IPR001452; SH3.
DR InterPro: IPR001245; Tyr_kinase.
DR InterPro: IPR008266; Tyr_kinase_AS.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF00017; SH2; 1.
DR Pfam: PF00018; SH3; 1.
DR PRINTS: PR00401; SH2DOMAIN.
DR PRINTS: PR00452; SH3DOMAIN.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Prot_kinase; 1.
DR ProDom: PD000093; SH2; 1.
DR ProDom: PD000066; SH3; 1.
DR SMART: SM00252; SH2; 1.

```

```
DR SMART; SMO0326; SH3; 1.
DR SMART; SMO0219; TYRKC; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00001; SH2; 1.
DR PROSITE; PS00002; SH3; 1.
DR PROSITE; PS00003; SH3; 1.
DR TRANSFERASE; Tyrosine-protein kinase; ATP-binding; SH2 domain;
DR SH3 domain; Phosphorylation.
FT DOMAIN 11 72
FT DOMAIN 78 170
FT DOMAIN 191 445
FT NP_BIND 197 205
FT BINDING 219 219
FT ACT_SITE 312 312
FT ACT_SITE 342 342
FT MOD_RES 447 447
FT MOD_RES 451 451
SQ SEQUENCE 451 AA; 51972 MW; 8F5EC26329498DEC CRC64;

Query Match 53.4%; Score 39; DB 1; Length 451;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 5 LWWNA 9
Db 43 LWWNA 47

RESULT 25
NUOM BUCAP STANDARD; PRT; 501 AA.
AC Q8K9X6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE NADH-quinone oxidoreductase chain M (EC 1.6.99.5) (NADH dehydrogenase
DE 1, chain M) (NDH-1, chain M).
GN NUOM OR BUSG158.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22084549; PubMed=12089438;
RA Tamás I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
RA Wernegreen J.J., Sandstroem J.P., Moran N.A., Andersson S.G.E.;
RT "50 million years of genomic stasis in endosymbiotic bacteria.";
RL Science 236:2376-2379(2002).
CC -!- FUNCTION: NDH-1 shuttles electrons from NADH, via FMN and iron-
CC sulfur (Fe-S) centers, to quinones in the respiratory chain.
CC Couples the redox reaction to proton translocation (for every two
CC electrons transferred, four hydrogen ions are translocated across
CC the cytoplasmic membrane), and thus conserves the redox energy in
CC a proton gradient (By similarity).
CC -!- CATALYTIC ACTIVITY: NADH + quinone = NAD(+) + quinol.
CC -!- SUBUNIT: Composed of 13 different subunits. Subunits nuoA, H, J,
CC K, L, M, N constitute the membrane sector of the complex (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: TO POLYPEPTIDE 4 OF THE NADH-UBIQUINOL OXIDOREDUCTASE
CC OF CHLOROPLASTS OR MITOCHONDRIA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; AE014091; AA067726.1; --
DR InterPro; IPR003918; NADHud_oxred4.
```

```
DR InterPro; IPR001750; Oxidored_q1.
DR Pfam; PF00361; Oxidored_q1; 1.
DR PRINTS; PR01437; NUOXDRDTASE4.
DR Oxidoreductase; NAD; Quinone; Transmembrane; Complete proteome.
FT TRANSMEM 1 21
FT TRANSMEM 30 50
FT TRANSMEM 86 106
FT TRANSMEM 115 134
FT TRANSMEM 139 156
FT TRANSMEM 174 194
FT TRANSMEM 222 242
FT TRANSMEM 252 272
FT TRANSMEM 286 306
FT TRANSMEM 314 331
FT TRANSMEM 341 363
FT TRANSMEM 374 394
FT TRANSMEM 420 440
FT TRANSMEM 459 479
SQ SEQUENCE 501 AA; 57884 MW; 2218E31ED3EFFDE2 CRC64;

Query Match 53.4%; Score 39; DB 1; Length 501;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 WTNINW 7
Db 378 WTNINW 383

RESULT 26
OPUD BACSU STANDARD; PRT; 512 AA.
AC P54417;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glycine betaine transporter opuD.
GN OPUD OR BSU30070.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=168 / JH642;
RX MEDLINE=96359357; PubMed=8752321;
RA Kappes R., Kempf B., Bremer E.;
RT "Three transport systems for the osmoprotectant glycine betaine
RT operate in Bacillus subtilis: characterization of OpuD.";
RL J. Bacteriol. 178:5071-5079(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=98048467; PubMed=9387221;
RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RT "Sequencing and functional annotation of the Bacillus subtilis genes
RT in the 200 kb rnb-dnaB region.";
RL Microbiology 143:3431-3441(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunz F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bartoro M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Etikian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Gusepki G., Guy B.J., Haga K., Halect J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Iulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
```

RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapina A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medique C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly T.M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Forwollik S., Prescott A.M.,
RA Priesan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Serró S.J., Serró P., Shin B.S., Solido B.,
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takenaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viñari A., Wambitt R., Wedler K., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*".
RT Nature 390:249-256(1997).
CC -!- FUNCTION: High-affinity uptake of glycine betaine.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to the bccT transporter (TC 2.A.15) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U50082; AAC43468.1; -.
DR EMBL: AF008220; AAC00408.1; -.
DR EMBL: Z99119; CAB14985.1; -.
DR FIC: G69670; G69670.
DR Subtilist; BG11958; op:3.
DR InterPro: IPR000060; BCC_T transporter.
DR Pfam: PF02028; BCC_T; 1.
DR ProDom: PD01011; BCC_T transporter; 1.
DR TIGRFAMs: TIGR00842; bccT; 1.
DR PROSITE: PS01303; BCC_T; 1.
KW Transport; Transmembrane; Complete proteome.
FT TRANSMEM 5 25 POTENTIAL.
FT TRANSMEM 45 65 POTENTIAL.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 135 155 POTENTIAL.
FT TRANSMEM 186 206 POTENTIAL.
FT TRANSMEM 222 242 POTENTIAL.
FT TRANSMEM 257 277 POTENTIAL.
FT TRANSMEM 312 332 POTENTIAL.
FT TRANSMEM 343 363 POTENTIAL.
FT TRANSMEM 395 415 POTENTIAL.
FT TRANSMEM 441 461 POTENTIAL.
FT TRANSMEM 464 484 POTENTIAL.
SQ SEQUENCE 512 AA; 56122 MW; FID8B1E8EE30F1PZ CRC64;
Query Match 53.4%; Score 39; DB 1; Length 512;
Best Local Similarity 42.9%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 6; Gaps 1;
QY 1 EWTNIN-----WW 8
Db 307 EWINSWTIFWAWW 320
RESULT 27
VNL1 DROME STANDARD; PRT; 555 AA.
AC Q3NF1; Q3NF1; 16-OCT-2003 (Rel. 40, Created)
DT 16-OCT-2003 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vanin-like protein 1 precursor.

GN CG32754/CG3648.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI TaxID=7227;
RP [1]_SEQUENCE FROM N.A.
RX MEDLINE=99432008; PubMed=10501839;
RT Granjeaud S., Naquet P., Galland F.;
RT "An ESTs description of the new Vanin gene family conserved from fly
RT to human".
RL Immunogenetics 49:964-972(1999).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.P.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Watanabe K., Rogers Y.-H.C., Blazek R.G., Champagne M., Pfeiffer B.D.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Fogle C., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan X., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*".
RL Science 287:2185-2195(2000).
RN [3]
RN REVISIONS.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celisner S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.B.;
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
RT systematic review".
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [4]
RN SEQUENCE FROM N.A.
RC STRAIN=Berkeley; TISSUE=Larva, and Pupae;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,

George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
 Rubin G.M., Celniker S.E.; cDNA resource.";
 "A Drosophila full-length cDNA resource.";
 Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
 CC -!- FUNCTION: Probable hydrolase.
 CC -!- TISSUE SPECIFICITY: Expressed in larvae and early pupae.
 CC -!- SIMILARITY: Belongs to the CN hydrolase family. BTB/VNN subfamily.
 CC -!- SIMILARITY: Contains 1 CN hydrolase domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AJ276261; CAB77020.1; -;
 CC EMBL; AB003436; AAN09161.1; -;
 CC EMBL; AY052034; AAK93458.1; -;
 CC FlyBase; F3gn040069; vanin-like.
 CC GO; GO:0013898; C:extrinsic to membrane; NAS.
 CC GO; GO:0016787; F:hydrolase activity; NAS.
 CC InterPro; IPR003010; Ntlase/CNhydase.
 CC Pfam; PF00795; CN hydrolase; 1.
 CC PROSITE; PS0263; CN HYDROLASE; 1.
 CC Hydrolase; Signal; Glycoprotein.
 CC SIGNAL 1 22 POTENTIAL.
 CC CHAIN 23 558 VANIN-LIKE PROTEIN 1.
 CC DOMAIN 33 344 CN HYDROLASE.
 CC FT CARBOHYD 65 65 (POTENTIAL).
 CC FT CARBOHYD 103 103 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 120 120 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 128 128 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 180 180 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 354 354 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC FT CARBOHYD 379 379 N-LINKED (GLCNAC. .) (POTENTIAL).
 CC SEQUENCE 558 AA; 62341 MW; 7920621923BB7779 CRC64;
 Query Match 53.4%; Score 39; DB 1; Length 558;
 Best Local Similarity 66.7%; Pred. No. 1.7e-02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Qy 3 TNINWW 8
 Db 2 SNTNWW 7
 RESULT 28
 ID AD10 BOVIN STANDARD; PRT; 748 AA.
 AC Q10741;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE 10-OCT-2003 (Rel. 42, Last annotation update)
 DE ADAM 10 precursor (EC 3.4.24.81) (A disintegrin and metalloproteinase
 DE domain 10) (Mammalian disintegrin-metalloprotease) (Myelin-associated
 DE metalloproteinase) (Kuzbanian protein homolog).
 GN ADAM10 OR YAM.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=96276398; PubMed=8694785;
 RA Howard L., Lu X., Mitchell S., Griffiths S., Glynn P.;
 RA "Molecular cloning of YAM: a catalytically active mammalian
 FT disintegrin-metalloprotease expressed in various cell types.";
 RL Biochem. J. 317:45-50(1996).
 RN [2]

SEQUENCE FROM N.A.; SEQUENCE OF 214-233, FUNCTION, AND MUTAGENESIS OF
 GLU-384.
 RP MEDLINE=99199286; PubMed=10097139;
 RX Lammich S., Kojro E., Postina R., Gilbert S., Pfeiffer R.,
 RA Jasionowski M., Haass C., Jahnholz F.;
 RT "Constitutive and regulated alpha-secretase cleavage of Alzheimer's
 RT amyloid precursor protein by a disintegrin metalloprotease.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:3922-3927(1999).
 CC -!- FUNCTION: Cleaves the membrane-bound precursor of TNF-alpha to its
 CC mature soluble form. Responsible for the proteolytic release of
 CC several other cell-surface proteins, including heparin binding
 CC epidermal growth-like factor, ephrin-A2 and for constitutive and
 CC regulated alpha-secretase cleavage of amyloid precursor protein
 CC (APP). Contributes to the normal cleavage of the cellular prion
 CC protein. Involved in the cleavage of the adhesion molecule L1 at
 CC the cell surface and in released membrane vesicles, suggesting a
 CC vesicle-based protease activity. Controls also the proteolytic
 CC processing of Notch (By similarity).
 CC -!- CATALYTIC ACTIVITY: Endopeptidase of broad specificity.
 CC -!- COFACTOR: Binds 1 zinc ion (By similarity).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. The proteolytically
 CC active form is localized mainly in the plasma membrane whereas
 CC the proenzyme is found intracellularly in the Golgi.
 CC -!- TISSUE SPECIFICITY: Expressed at low level in kidney, spleen,
 CC lung, adrenal, heart and peripheral nerve.
 CC -!- INDUCTION: By interleukin-1 alpha in nasal cartilage.
 CC -!- PTM: The precursor is cleaved by a furin endopeptidase (By
 CC similarity).
 CC -!- SIMILARITY: Belongs to peptidase family M12B.
 CC -!- SIMILARITY: Contains 1 disintegrin domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z21961; CAA79973.1; -;
 CC FIR; S66129; S66129.
 CC HSP; P18619; LFVL.
 CC InterPro; IPR001762; Disintegrin.
 CC InterPro; IPR001818; Pept_M10A_M12B.
 CC InterPro; IPR006025; Pept_M_Zn_BS.
 CC InterPro; IPR001590; Peptidase_M12B.
 CC Pfam; PF00200; disintegrin; 1.
 CC Pfam; PF01421; Reprolysin; 1.
 CC SMART; SM00050; DISIN; 1.
 CC PROSITE; PS00215; ADAM_MEPPO; 1.
 CC PROSITE; PS00546; CYSTEINE_SWITCH; FALSE NEG.
 CC PROSITE; PS00427; DISINTEGRIN_1; FALSE NEG.
 CC PROSITE; PS00214; DISINTEGRIN_2; 1.
 CC PROSITE; PS00142; ZINC_PROTEASE; 1.
 CC Hydrolase; Metalloprotease; Zinc; Signal; Transmembrane; SH3-binding;
 CC Zymogen; Glycoprotein.
 CC SIGNAL 1 19 POTENTIAL.
 CC FT PROPEP 20 213 BY SIMILARITY.
 CC FT CHAIN 214 748 ADAM 10.
 CC FT DOMAIN 20 672 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 673 696 POTENTIAL.
 CC FT DOMAIN 697 748 CYTOPLASMIC (POTENTIAL).
 CC FT DOMAIN 214 456 METALLOPROTEASE.
 CC FT DOMAIN 457 554 DISINTEGRIN-LIKE.
 CC FT DOMAIN 555 595 CYS-RICH.
 CC FT SITE 173 173 CYSTEINE SWITCH (POTENTIAL).
 CC FT SITE 708 715 SH3-BINDING (POTENTIAL).
 CC FT SITE 722 728 SH3-BINDING (POTENTIAL).
 CC FT METAL 383 383 ZINC (CATALYTIC) (BY SIMILARITY).
 CC FT ACT SITE 384 384 ZINC (CATALYTIC) (BY SIMILARITY).
 CC FT METAL 387 387 ZINC (CATALYTIC) (BY SIMILARITY).
 CC FT METAL 393 393 ZINC (CATALYTIC) (BY SIMILARITY).
 CC FT DISULFID 222 313 BY SIMILARITY.

```

FT DISULFID 344 451 BY SIMILARITY.
FT DISULFID 399 435 BY SIMILARITY.
FT CARBOHYD 267 267 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 278 278 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 439 439 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 551 551 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT MUTAGEN 384 384 E->A: DECREASED STIMULATED AND
SQ SEQUENCE 748 AA; 84198 MW; 202E29830611F9E1 CRC64;

Query Match 53.4%; Score 39; DB 1; Length 748;
Best Local Similarity 55.6%; Pred. No. 2.3e+02;
Matches 5; Conservative -; Mismatches 3; Indels 0; Gaps 0;

QY 1 EWTNIWWNA 9
Db 672 EWIVAYWNA 680

RESULT 29
ABC6 HUMAN STANDARD; PRT; 842 AA.
AC Q9NP58; 075542;
AT 16-OCT-2001 (Rel. 40, Create)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE ATP-binding cassette, sub-family B, member 6, mitochondrial precursor
DE (Mitochondrial ABC transporter 3) (Mc-ABC transporter 3) (ABC
DE transporter umat).
GN ABCB6 OR MTABC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
EX MEDLINE=20298824; PubMed=10837493;
RA Mitsuhashi N., Miki T., Senbongi H., Yokoi N., Yano H., Miyazaki M.,
RA Nakajima N., Iwakaga T., Yokoyama Y., Shibata T., Seino S.;
RT "MTABC3, a novel mitochondrial ATP-binding cassette protein involved
RT in iron homeostasis";
EL J. Biol. Chem. 275:17536-17540 (2000).
[2]
RP SEQUENCE FROM N.A.
RA Hirsch-Ernst K.I., Schaefer A., Ernst B.P., Schmitz-Salue C.,
RA Awuah D., Kahl G.F.;
RT "Subcellular localization of the ABC transporter umat.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBSJ databases.
[3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.K., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan B., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ussdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., McIlhenny S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S.E., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shcherbenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzyzinski M.I., Skalska U., Smalley D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[4]

```

```

RP SEQUENCE OF 332-842 FROM N.A.
RC TISSUE=Brain;
RA Yu W., Gibbs R.A.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBSJ databases.
CC -!- FUNCTION: Could be involved in the transport of heme from the
CC mitochondria to the cytosol.
CC -!- SUBUNIT: Homodimer or heterodimer (Potential).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane (Potential).
CC -!- TISSUE SPECIFICITY: Widely expressed. Highest expression in heart
CC and skeletal muscles.
CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ289233; CAB95766.2; -
CC EMBL; AF076775; BAA75107.1; -
CC EMBL; AB039371; BAA96733.1; -
CC EMBL; AB039353; BAA96733.1; JOINED
CC EMBL; AB039354; BAA96733.1; JOINED
CC EMBL; AB039355; BAA96733.1; JOINED
CC EMBL; AB039356; BAA96733.1; JOINED
CC EMBL; AB039357; BAA96733.1; JOINED
CC EMBL; AB039358; BAA96733.1; JOINED
CC EMBL; AB039359; BAA96733.1; JOINED
CC EMBL; AB039360; BAA96733.1; JOINED
CC EMBL; AB039361; BAA96733.1; JOINED
CC EMBL; AB039362; BAA96733.1; JOINED
CC EMBL; AB039363; BAA96733.1; JOINED
CC EMBL; AB039364; BAA96733.1; JOINED
CC EMBL; AB039365; BAA96733.1; JOINED
CC EMBL; AB039366; BAA96733.1; JOINED
CC EMBL; AB039367; BAA96733.1; JOINED
CC EMBL; AB039368; BAA96733.1; JOINED
CC EMBL; AB039369; BAA96733.1; JOINED
CC EMBL; AB039370; BAA96733.1; JOINED
CC EMBL; BC000559; AAH00559.1; -
CC EMBL; AF070598; AAC28653.1; -
CC HSSP; P13569; INED.
CC Genew; HGNC:47; ABCB6.
CC MIM; 605452; -
CC InterPro; IPR003593; AAA_ATPase.
CC InterPro; IPR001140; ABC_TM_transp.
CC InterPro; IPR003439; ABC_transporter.
CC Pfam; PF00664; ABC_membrane; 1.
CC Pfam; PF00005; ABC_tran; 1.
CC ProDom; P000006; ABC_transporter; 1.
CC SMART; SM00382; AAA; 1.
CC PROSITE; PS00929; ABC_TM1F; 1.
CC PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
CC PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
CC ATP-binding; Transmembrane; Transport; Mitochondrion; Inner membrane;
CC Transmembrane; Transport; Mitochondrion; Inner membrane;
CC TRANSIT 1 842 ? MITOCHONDRION (POTENTIAL).
CC CHAIN 1 842 ? ATP-BINDING CASSETTE, SUB-FAMILY B,
CC MEMBER 6
CC TRANSMEM 27 47 POTENTIAL.
CC TRANSMEM 107 127 POTENTIAL.
CC TRANSMEM 148 168 POTENTIAL.
CC TRANSMEM 186 206 POTENTIAL.
CC TRANSMEM 376 396 POTENTIAL.
CC TRANSMEM 398 418 POTENTIAL.
CC TRANSMEM 502 522 POTENTIAL.
CC NP_BIND 623 630 ATP (POTENTIAL).
CC SEQUENCE 842 AA; 93885 MW; E63A7D59DCESB9ED CRC64;

Query Match 53.4%; Score 39; DB 1; Length 842;

```

Best Local Similarity 80.0%; Pred. No. 2.5e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 WWAK 10
DB 173 WWAR 177

RESULT 30

YNS3_YEAST
ID YNS3_YEAST STANDARD; PRT; 904 AA.
AC P42842;
DC 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical: 102.3 kDa protein in DAL82-RFA2 intergenic region.
GN YNL313C OR N0364.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID:4932;
RY [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / FY1676;
RX MEDLINE=96076632; PubMed=7502583;
RA Maftahi M., Nicaud J.-M., Levesque H., Gaillardin C.;
RT "Sequencing analysis of a 24.7 kb fragment of yeast chromosome XIV
family and ten new open reading frames.";
RL Yeast 11:1077-1085 (1995).
RN [2]
RP SEQUENCE OF 1-333 FROM N.A.
RA Maurer C.T.C., Urbanus J.H.M., Planta R.J.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO C.ELEGANS T20B12.1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; 246259; CAA86383.1; -;
DR EMBL; 271588; CAA96242.1; -;
DR EMBL; 271589; CAA96243.1; -;
DR F01; S51299; S51299.
DR Germline; 143319; -;
DR SGD; S0005257; YNL313C.
DR InterPro; IPR008941; TPR-like.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 3.
KW Hypothetical protein.
SQ SEQUENCE 904 AA; 102336 MW; 7170398FA95A0BD8 CRC64;

Query Match 53.4%; Score 39; DB 1; Length 904;
Best Local Similarity 80.0%; Pred. No. 2.7e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 WWAK 10
DB 195 WWAR 199

RESULT 31

MGR4_HUMAN
ID MGR4_HUMAN STANDARD; PRT; 912 AA.
AC Q14833;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Metabotropic glutamate receptor 4 precursor (mGluR4).

GN GRM4 OR GPRCID OR MGLUR4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96346635; PubMed=8738157;
RA Makoff A., Lelchuk R., Oxer M., Harrington K., Emson P.;
RT "Molecular characterization and localization of human metabotropic
glutamate receptor type 4.";
RL Brain Res. Mol. Brain Res. 37:239-248 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98141892; PubMed=9473604;
RA Wu S., Wright R.A., Rockey P.K., Burchett S.G., Arnold J.S.,
RA Roetack P.R. Jr., Johnson B.G., Schoepp D.D., Belagaje R.M.;
RT "Group III human metabotropic glutamate receptors 4, 7 and 8:
molecular cloning, functional expression, and comparison of
pharmacological properties in RGT cells.";
RL Brain Res. Mol. Brain Res. 53:88-97 (1998).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95342351; PubMed=7617140;
RA Flor P.J., Lukic S., Rueegg D., Leonhardt T., Knoepfel T., Kuhn R.;
RT "Molecular cloning, functional expression and pharmacological
characterization of the human metabotropic glutamate receptor type
4.";
RL Neuropharmacology 34:149-155 (1995).

CC VARIANT ILE-797.
RX MEDLINE=21416233; PubMed=11525421;
RA Ohtsuki T., Toru M., Arinami T.;
RT "Mutation screening of the metabotropic glutamate receptor mGluR4
(GRM4) gene in patients with schizophrenia.";
RL Psychiatr. Genet. 11:79-83 (2001).
CC -!- FUNCTION: Receptor for glutamate. The activity of this receptor
is mediated by a G-protein that inhibits adenylyl cyclase
activity.
CC -!- SUBUNIT: Interacts with PRKCAPB (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- TISSUE SPECIFICITY: Strongly expressed in the cerebellum.
CC Expressed at low levels in hippocampus, hypothalamus and
thalamus. No expression detected in liver.
CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
STRONGEST, TO MGLUR6.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; X80818; CAA56784.1; -;
DR EMBL; U92457; AAB51762.1; -;
DR Genew; HGNC:4596; GRM4.
DR MIM; 604100;
DR GO; GO:0005887; C: integral to plasma membrane; TAS.
DR GO; GO:0008067; P: metabotropic glutamate, GABA-B-like recepto. . . ; TAS.
DR GO; GO:0007194; P: negative regulation of adenylyl cyclase ac. . . ; TAS.
DR GO; GO:0007268; P: synaptic transmission; TAS.
DR InterPro; IPR001828; ANF receptor.
DR InterPro; IPR000337; GPCR_Mgr.
DR Pfam; PF00003; 7tm_3; 1.
DR Pfam; PF01094; ANF_receptor; 1.
DR PRINTS; PR00248; GPCRMR.
DR PROSITE; PS00979; G_PROTEIN_RECEP_F3_1; 1.
DR PROSITE; PS00980; G_PROTEIN_RECEP_F3_2; 1.
DR PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; 1.

DR PROSITE; PS0259; G_PROTEIN_RECP_F3_4; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
 KX Multigene family; Polymorphism.
 FT SIGNAL 1 32
 FT CHAIN 33 912
 FT DOMAIN 33 912
 FT TRANSMEM 588 610
 FT DOMAIN 611 624
 FT TRANSMEM 625 645
 FT DOMAIN 646 656
 FT TRANSMEM 657 675
 FT DOMAIN 676 699
 FT TRANSMEM 700 720
 FT DOMAIN 721 750
 FT TRANSMEM 751 772
 FT DOMAIN 773 785
 FT TRANSMEM 786 808
 FT DOMAIN 809 821
 FT TRANSMEM 822 847
 FT DOMAIN 848 912
 FT CARBOHYD 98 98
 FT CARBOHYD 301 301
 FT CARBOHYD 454 454
 FT CARBOHYD 484 484
 FT CARBOHYD 569 569
 FT VARIANT 797 797
 SQ SEQUENCE 912 AA; 101867 MW; 4A2F36E63A2EAF5A CRC64;
 Query Match 53.4%; Score 39; DB 1; Length 912;
 Best Local Similarity 80.0%; Pred. No. 2.8e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 6 WWWAK 10
 DB 9 WWWAR 13
 RESULT 32
 MGR4 RAT STANDARD; PRT; 912 AA.
 AC P31423;
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Metabotropic glutamate receptor 4 precursor (mglur4).
 GN GRM4 OR GPRC4D OR MGLUR4.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CX NCBI_TaxID=10116;
 EN [1]
 SEQUENCE FROM N.A.
 RP TISSUE=Brain;
 RC MEDLINE=92110002; PubMed=1309649;
 RX Tarabe Y., Masu M., Ishii T., Shigemoto R., Nakanishi S.;
 RA "A family of metabotropic glutamate receptors.";
 RT Neuron 8:169-179(1992).
 RL [2]
 SEQUENCE FROM N.A.
 RP TISSUE=Brain;
 RC MEDLINE=9332693; PubMed=8338667;
 RX O'Hara P.J., Sheppard P.O., Thøgersen H., Venezia D., Haldeman B.A.,
 RA McGrane V., Houamed K.M., Thomsen C., Gilbert T.L., Mulvihill E.R.;
 RT "The ligand-binding domain in metabotropic glutamate receptors is
 related to bacterial periplasmic binding proteins.";
 RL Neuron 11:41-52(1993).
 RN [3]
 INTERACTION WITH PRKCBP.
 RP MEDLINE=20571357; PubMed=11122333;
 RX El Far O., Airas J., Wischmeyer E., Nehring R.B., Karschin A.,
 RA Betz H.;
 RT "Interaction of the C-terminal tail region of the metabotropic

RT glutamate receptor 7 with the protein kinase C substrate PICK1.";
 RL Eur. J. Neurosci. 12:4215-4221(2000).
 CC -!- FUNCTION: Receptor for glutamate. The activity of this receptor
 CC is mediated by a G-protein that inhibits adenylate cyclase
 CC activity.
 CC -!- SUBUNIT: Interacts with PRKCBP.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: Is widely distributed in the CNS. Predominant
 CC expression is seen in the granule cells of the cerebellum.
 CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
 CC STRONGEST, TO MGLUR6.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M92077; -; NOT ANNOTATED_CDS.
 CC EMBL; M90518; AAA93190.1; -;
 CC PIR; JH0563; JH0563.
 CC InterPro; IPR001828; ANF receptor.
 CC InterPro; IPR000337; GPCR_Mgr.
 CC Pfam; PF00003; 7tm_3; 1.
 CC Pfam; PF01094; ANF_receptor; 1.
 CC PRINTS; PR00248; GPCRMR.
 CC PROSITE; PS00979; G_PROTEIN_RECP_F3_1; 1.
 CC PROSITE; PS00980; G_PROTEIN_RECP_F3_2; 1.
 CC PROSITE; PS00981; G_PROTEIN_RECP_F3_3; 1.
 CC PROSITE; PS00259; G_PROTEIN_RECP_F3_4; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
 KW Multigene family.
 FT SIGNAL 1 32
 FT CHAIN 33 912
 FT DOMAIN 33 587
 FT TRANSMEM 588 610
 FT DOMAIN 611 624
 FT TRANSMEM 625 643
 FT DOMAIN 646 656
 FT TRANSMEM 657 675
 FT DOMAIN 676 699
 FT TRANSMEM 700 720
 FT DOMAIN 721 750
 FT TRANSMEM 751 772
 FT DOMAIN 773 785
 FT TRANSMEM 786 808
 FT DOMAIN 809 821
 FT TRANSMEM 822 847
 FT DOMAIN 848 912
 FT CARBOHYD 98 98
 FT CARBOHYD 301 301
 FT CARBOHYD 454 454
 FT CARBOHYD 484 484
 FT CARBOHYD 569 569
 FT CONFLICT 124 124
 SQ SEQUENCE 912 AA; 101818 MW; 336433EF19B4B577 CRC64;
 Query Match 53.4%; Score 39; DB 1; Length 912;
 Best Local Similarity 80.0%; Pred. No. 2.8e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 6 WWWAK 10
 DB 9 WWWAR 13
 RESULT 33
 ASPI_MOUSE
 ID ASPI_MOUSE STANDARD; PRT; 1087 AA.
 AC Q62415;
 DT 15-JUL-1998 (Rel. 36, Created)

Dfam: PF00018; SH3; 1.
DR PRINTS; PR06452; SH3DOMAIN.
DR ProDom; PD00066; SH3; 1.
DR SMART; SM00248; ANK; 2.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50088; ANK_REPEAT; 2.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50002; SH3; 1.
KW Apoptosis; Nuclear protein; Repeat; ANK repeat; SH3 domain.
FT REPEAT 917 949 ANK 1.
FT REPEAT 950 982 ANK 2.
FT DOMAIN 1016 1078 SH3.
FT DOMAIN 132 292 GLN-RICH.
FT DOMAIN 446 857 PRO-RICH.
SQ SEQUENCE 1087 AA; 119169 MW; 8B3E9CC4B2339F13 CRC64;

Query Match 53.4%; Score 39; DB 1; Length 1087;
Best Local Similarity 80.0%; Pred. No. 3.3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps

QY 6 WWAK 10
|||||
Db 1056 WWAR 1060

RESULT 34
ASP2_MOUSE
ID ASP2_MOUSE STANDARD; PRT; 1088 AA.
AC Q8CG79; O8K2L5;
AD 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Apoptosis stimulating of p53 protein 2 (tumor suppressor p53-binding protein 2).
DE TRP5BP2 OR ASPP2.
OS Mus musculus (Mouse).
GN Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxId=10090;
[1]
RP SEQUENCE FROM N.A.
RN STRAIN=FVE/N; TISSUE=Breast tumor;
RC MEDLINE=22388257; PubMed=12471932;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Bhat N.K., Hopkins R.E., Jordan H., Moore T., Max S.I., Wang J.J., Haieff F., Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E., Brownstein M.J., Udwin T.B., Toshiyuki S., Carrinci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W., Villalon D.K., Munzly D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Bladesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalus D.E., Scherch A., Schein J.E., Jones S.J.M., Marra M.A.; "Generation and initial analysis of more than 15,000 full-length human transcripts and mouse cDNA sequences."

-!- FUNCTION: Regulator that plays a central role in regulation of apoptosis and cell growth via its interactions. Regulates TP53 by enhancing the DNA binding and transactivation function of TP53 on the promoters of proapoptotic genes in vivo. Inhibits the ability of APPBP1 to conjugate NEBDP8 to CUL1, and thereby decreases APPBP1 ability to induce apoptosis. Impedes cell cycle progression at G2/M (By similarity).

-!- SUBUNIT: Binds to the central domain of TP53 as well as to BCL2. Interacts with protein phosphatase 1. Interacts with RELA NF-kappa-B subunit. This interaction probably prevents the activation of NF-kappa-B.

of apoptosis, possibly by preventing its interaction with TP53.
 Interacts with APPB1 (By similarity).
 SUBCELLULAR LOCATION: Predominantly cytoplasmic; perinuclear region. Some small fraction is nuclear (By similarity).
 DOMAIN: The ankyrin repeats and the SH3 domain are required for a specific interactions with TP53 (By similarity).
 SIMILARITY: Belongs to the ASPP family.
 SIMILARITY: Contains 2 ANK repeats.
 SIMILARITY: Contains 1 SH3 domain.
 CAUTION: Ref 1 (AAH30894) sequence differs from that shown due to a frameshift in position 23.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; BC030894; AAH30894.1; ALT FRAME.
 EMBL; BC042874; AAH42874.1; ALT_INIT.
 PIR; P0551; P0551.
 MGI; MGI:2138319; Trp53bp2.
 InterPro; IPR002110; ANK.
 InterPro; IPR001452; SH3.
 Pfam; PF00023; ank; 2.
 Pfam; PF00018; SH3; 1.
 ProDom; PD000066; SH3; 1.
 SMART; SM00248; ANK; 2.
 SMART; SM00326; SH3; 1.
 PROSITE; PS50088; ANK_REPEAT; 2.
 PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 PROSITE; PS50002; SH3; 1.
 Apoptosis; Cell cycle; SH3 domain; SH3-binding; Repeat; ANK repeat.
 REPEAT 918 950 ANK 1.
 REPEAT 951 983 ANK 2.
 DOMAIN 1017 -1079 SH3.
 DOMAIN 292 308 INTERACTION WITH APPB1 (BY SIMILARITY).
 DOMAIN 92 133 GLN-RICH.
 SITE 826 835 SH3-BINDING (POTENTIAL).
 SEQUENCE 1088 AA; 120731 MW; 1023B229099BF3EC CRC64;

Query Match 53.4%; Score 39; DB 1; Length 1088;
 Best Local Similarity 80.0%; Pred. No. 3.3e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 WWAK 13
 Db 1057 WWAR 1061

RESULT 35
 ASPI HUMAN
 ID ASPI HUMAN STANDARD; PRT; 1050 AA.
 AC Q36KQ4; O94870;
 DT 13-OCT-2003 (Rel. 42, Created)
 DT 13-OCT-2003 (Rel. 42, Last sequence update)
 DE Apoptosis stimulating of p53 protein 1 (Protein phosphatase 1 regulatory subunit 13B).
 GN PP1R13B OR ASPP1 OR KIAA0771.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.; FUNCTION, SUBCELLULAR LOCATION, DISEASE, AND INTERACTION WITH TP53.
 RX MEDLINE=21541920; PubMed=11684014;
 RA Samuels-Lev Y., O'Connor D.J., Bergamaschi D., Trigliante G., Hsieh J.-K., Zhong S., Campargue I., Naumovski L., Crook T., Lu X.;
 RT "ASPP proteins specifically stimulate the apoptotic function of p53.";

Mol. Cell 8:791-794 (2001).
 [2]
 RN SEQUENCE FROM N.A.
 RX PubMed=12508121;
 RA Heilig R., Eckenberg R., Petit J.-L., Fonknechten N., Da Silva C., Cattolico L., Levy M., Barbe V., De Berardinis V., Ureca-Vidal A., Pelletier E., Vico M., Anthouard V., Rowen L., Madan A., Qin S., Sun H., Du H., Pepin K., Artiguenave F., Robert C., Cruaud C., Bruels T., Jallou O., Friedlander L., Samson G., Brottier P., Cure S., Segurens B., Aniere F., Samain S., Crespeau H., Abbasi N., Alich N., Boscus D., Dickhoff R., Dors M., Dubois I., Friedman C., Gouvenoux M., James R., Madan A., Maire-Estrada B., Mangenot C., Martins N., Menard M., Oztas S., Ratcliffe A., Shaffer T., Trask B., Vacherie B., Bellemere C., Belser C., Besnard-Gonnet M., Bartol-Mavel D., Bontard M., Briez-Silla S., Combette S., Bufosse-Laurent V., Ferron C., Lechaplais C., Louesse C., Muselet D., Magdelenat G., Pateau E., Petit E., Sirvain-Trukiewicz P., Trybou A., Vega-Czarny N., Bataille E., Bluet E., Bordelais I., Dubois M., Dumont C., Guerin T., Haffray S., Hamadi R., Muanga J., Pellouin V., Robert D., Wunderle E., Gauguier G., Roy A., Sainte-Marthe L., Verdier J., Verdier-Discala C., Hillier L.W., Fulton L., McPherson J., Matsuda F., Wilson R., Scarpelli C., Gyapay G., Wincker P., Saurin W., Quetier F., Waterston R., Hood L., Weissbach J.;
 RT "The DNA sequence and analysis of human chromosome 14.";
 Nature 421:601-607 (2003).
 [3]
 RN SEQUENCE OF 143-1090 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=99087487; PubMed=9872452;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XI. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";
 DNA Res. 5:277-286 (1998).
 CC -!- FUNCTION: Regulator that plays a central role in regulation of apoptosis via its interaction with p53/TP53. Regulates TP53 by enhancing the DNA binding and transactivation function of TP53 on the promoters of proapoptotic genes in vivo.
 CC -!- SUBUNIT: Interacts with TP53.
 CC -!- SUBCELLULAR LOCATION: Predominantly cytoplasmic. Some fraction is nuclear.
 CC -!- DOMAIN: The ankyrin repeats and the SH3 domain are required for a specific interactions with TP53.
 CC -!- DISEASE: Defects in PP1R13B may be a cause of breast cancers. It is overexpressed in many patients suffering from breast carcinomas and expressing a wild-type TP53 protein.
 CC -!- MISCELLANEOUS: In contrast to its official gene name, it is not a regulatory subunit of protein phosphatase 1. This name was given due to its similarity with a protein that binds to protein phosphatase 1.
 CC -!- SIMILARITY: Belongs to the ASPP family.
 CC -!- SIMILARITY: Contains 2 ANK repeats.
 CC -!- SIMILARITY: Contains 1 SH3 domain.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AJ318887; CAC83011.2; -
 EMBL; AL049840; -; NOT ANNOTATED_CDS.
 EMBL; AB018314; BAA34491.1; -
 GenBank; HGNC:14950; PP1R13B.
 MIM; 606455; -
 InterPro; IPR021110; ANK.
 InterPro; IPR001452; SH3.
 Pfam; PF00023; ank; 2.
 Pfam; PF00018; SH3; 1.
 PRINTS; PR00452; SH3DOMAIN.


```

FT STRAND      1061      1064
FT STRAND      1068      1068
FT TURN        1073      1074
FT STRAND      1075      1075
FT STRAND      1078      1078
FT STRAND      1083      1086

Query Match          53.4%; Score 39; DB 1; Length 1128;
Best Local Similarity 80.0%; Pred.No. 3.4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy       6 WWWAK 10
Db       1097 WWWAR 1101
|||||

RESULT 37
RCEL RHOSH
ID RCEL RHOSH STANDARD; PRT; 201 AA.
AC P02954; Q9RFB9;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Reaction center protein L chain (Photosynthetic reaction center L
DE subunit).
GN PUFL.
OS Rhodobacter sphaeroides (Rhodopseudomonas sphaeroides).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Rhodobacter.
OX NCBI_TaxID=1063;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=85063779; PubMed=6095283;
RA Williams J.C., Steiner L.A., Feher G., Simon M.I.;
RT "Primary structure of the L subunit of the reaction center from
RT Rhodopseudomonas sphaeroides.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:7303-7307(1984).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=Y;
RX MEDLINE=91129295; PubMed=2126457;
RA Arnoux B., Ducruix A., Astier C., Picaud M., Roth M., Reiss-Husson F.;
RT "Towards the understanding of the function of RB sphaeroides Y wild
RT type reaction center: gene cloning, protein and detergent structures
RT in the three-dimensional crystals.";
RL Biochimie 72:525-530(1990).
RN [3]
SEQUENCE FROM N.A.
RP STRAIN=ATCC 17023 / 2.4.1 / NCTB 8253 / DSM 158;
RX MEDLINE=20115911; PubMed=10648776;
RA Choudhary M., Kaplan S.;
RT "DNA sequence analysis of the photosynthesis region of Rhodobacter
RT sphaeroides 2.4.1.";
RL Nucleic Acids Res. 28:862-867(2000).
RN [4]
X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS).
RP MEDLINE=91242447; PubMed=2036404;
RX Chang C.-H., El-Kabbani O., Tiede D., Norris J., Schiffer M.;
RT "Structure of the membrane-bound protein photosynthetic reaction
RT center from Rhodobacter sphaeroides.";
RL Biochemistry 30:5352-5360(1991).
RN [5]
X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RP STRAIN=R-26;
RX MEDLINE=89042207; PubMed=3054889;
RA Allen J.P., Feher G., Yeates T.O., Komiya H., Rees D.C.;
RT "Structure of the reaction center from Rhodobacter sphaeroides R-26:
RT protein-cofactor (quinones and Fe+) interactions.";
Proc. Natl. Acad. Sci. U.S.A. 85:8487-8491(1988).
RN [6]
X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RP STRAIN=R-26;
RX MEDLINE=87317613; PubMed=2819866;

```

EA Allen J.P., Feher G., Yeates T.O., Komiya H., Rees D.C.;
RT "Structure of the reaction center from Rhodobacter sphaeroides R-26:
RL the protein subunits.";
RN Proc. Natl. Acad. Sci. U.S.A. 84:6162-6166(1987).
RP [7]
RX X-RAY CRYSTALLOGRAPHY (2.55 ANGSTROMS).
RA MEDLINE=96206889; PubMed=9537989;
RA McAuley-Hecht K.E., Pyfe P.K., Ridge J.P., Prince S.M., Hunter C.N.,
EA Isaacs N.W., Cogdell R.J., Jones M.R.;
RT "Structural studies of wild-type and mutant reaction centers from an
RT antenna-deficient strain of Rhodobacter sphaeroides: monitoring the
RT optical properties of the complex from bacterial cell to crystal.";
RL Biochemistry 37:4740-4753(1998).
RN [8]
RP TOPOLOGY.
RX MEDLINE=91250399; PubMed=1645718;
RA Yun C.H., Van Doren S.R., Crofts A.R., Gennis R.B.;
RT "The use of gene fusions to examine the membrane topology of the L-
RT subunit of the photosynthetic reaction center and of the cytochrome b
RT subunit of the bc1 complex from Rhodobacter sphaeroides.";
RL J. Biol. Chem. 266:10967-10973(1991).
CC -!- FUNCTION: THE REACTION CENTER IS A MEMBRANE-BOUND COMPLEX THAT
CC MEDIATES THE INITIAL PHOTOCHEMICAL EVENT IN THE ELECTRON TRANSFER
CC PROCESS OF PHOTOSYNTHESIS.
CC -!- SUBUNIT: REACTION CENTER IS COMPOSED OF FOUR BACTERIOCHLOROPHYLLS,
CC TWO BACTERIOPEOPHYTINS, TWO UBIQUINONES, ONE IRON, AND THREE
CC HIGHLY HYDROPHOBIC POLYPEPTIDE CHAINS (DESIGNATED L, M, AND H).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE REACTION CENTER L/M CHAINS / PSBA /
CC PSBD FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; M10206; AAA26177.1; -;
DR EMBL; X63404; CAA44999.1; -;
DR EMBL; AF195122; AAF24304.1; -;
DR PIR; S24212; WNRFLS.
DR PIR; T50760; T50760.
DR PDB; 2RCR; 15-JUL-93.
DR PDB; 4RCR; 31-OCT-93.
DR PDB; 1FSS; 30-APR-94.
DR PDB; 1PST; 30-APR-94.
DR PDB; 1PCR; 07-FEB-95.
DR PDB; 1AIG; 22-OCT-97.
DR PDB; 1AIJ; 22-OCT-97.
DR PDB; 1YST; 27-FEB-95.
DR PDB; 1MPS; 14-OCT-98.
DR PDB; 1DS8; 08-MAR-00.
DR PDB; 1DV3; 08-MAR-00.
DR PDB; 1DV6; 08-MAR-00.
DR PDB; 1E14; 23-AUG-00.
DR PDB; 1E6D; 18-JUL-03.
DR PDB; 1F6M; 18-APR-01.
DR PDB; 1FNP; 18-APR-01.
DR PDB; 1FNQ; 18-APR-01.
DR PDB; 1JGW; 05-SEP-01.
DR PDB; 1JGX; 05-SEP-01.
DR PDB; 1JGY; 05-SEP-01.
DR PDB; 1JGZ; 05-SEP-01.
DR PDB; 1JHO; 05-SEP-01.
DR PDB; 1KBY; 13-NOV-02.
DR PDB; 1L9B; 12-JUN-02.
DR PDB; 1L9J; 12-JUN-02.
DR PDB; 1M3X; 28-AUG-02.
DR PDB; 1QOV; 13-DEC-99.
DR InterPro; IPR005871; Photo L.
DR InterPro; IPR000484; Photo RC.

DR Pfam; PF00124; photoRC; 1.
DR PRINTS; PR00256; REACTCENTRE.
DR TIGRFA; TIGR01157; PUEL; 1.
DR PROSITE; PS00244; REACTION_CENTER; 1.
KW Electron transport; Photosynthesis; Reaction center;
KW Bacteriochlorophyll; Iron; Magnesium; Transmembrane; 3D-structure.
FT INIT MET 0 0
FT DOMAIN 1 31 CYTOPLASMIC.
FT TRANSMEM 32 55
FT DOMAIN 56 82 PERIPLASMIC.
FT TRANSMEM 83 111
FT DOMAIN 112 115 CYTOPLASMIC.
FT TRANSMEM 116 138
FT DOMAIN 139 170 PERIPLASMIC.
FT TRANSMEM 171 198
FT DOMAIN 199 224 CYTOPLASMIC.
FT TRANSMEM 225 250
FT DOMAIN 251 281 PERIPLASMIC.
FT METAL 153 153 MAGNESIUM (BACTERIOCHLOROPHYLL B AXIAL
LIGAND).
FT METAL 173 173 MAGNESIUM (BACTERIOCHLOROPHYLL B AXIAL
LIGAND).
FT METAL 190 190 IRON (NON HEME).
FT METAL 210 210 IRON (NON HEME).
FT BINDING 216 216 QUINONE B.
FT CONFLICT 54 54 V -> E (IN REF. 3).
FT CONFLICT 58 58 T -> S (IN REF. 3).
FT CONFLICT 65 65 S -> F (IN REF. 3).
FT STRAND 2 2
FT TURN 4 5
FT HELIX 7 9
FT TURN 19 22
FT STRAND 25 26
FT TURN 27 28
FT STRAND 29 30
FT HELIX 33 56
FT TURN 57 57
FT TURN 61 63
FT STRAND 66 66
FT HELIX 71 73
FT TURN 74 75
FT HELIX 80 82
FT TURN 83 83
FT HELIX 84 111
FT TURN 112 112
FT HELIX 116 132
FT TURN 133 133
FT HELIX 134 139
FT TURN 140 140
FT HELIX 142 144
FT STRAND 148 148
FT HELIX 150 163
FT TURN 164 164
FT HELIX 167 169
FT HELIX 171 198
FT TURN 202 203
FT HELIX 209 220
FT STRAND 222 222
FT HELIX 225 250
FT STRAND 251 251
FT TURN 252 254
FT STRAND 255 255
FT HELIX 259 261
FT TURN 262 263
FT HELIX 264 267
FT TURN 268 268
FT TURN 270 274
SQ SEQUENCE 281 AA; 31325 MW; 9D5C569B59ACFCA2 CRC64;

Query Match 52.7%; Score 38.5; DB 1; Length 281;
Best Local Similarity 45.5%; Pred. No. 1e+02;
Matches 5; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

```
QY 1 EWTN1W-WWAK 10
Db 258 QWVDWQWQWVK 268

Best Local Similarity 62.5%; Pred. No. 2e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 1 EWTN1W 8
Db 277 EWKS-WW 283

RESULT 39
SYM_AERPE STANDARD; PRT; 572 AA.
AC Q9YCY3;
DT 30-MAY-2000 (Rel. 39, Last sequence update);
DT 30-MAY-2000 (Rel. 39, Last sequence update);
DT 28-FEB-2003 (Rel. 41, Last annotation update);
DE Methionyl-tRNA synthetase (EC 6.1.1.10) (Methionine--tRNA ligase)
DE (Met-PS);
DE METG OR APEL129.
GN Aeropyrum pernix.
OS Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anka H., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushiida N., Ogunchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
CC -!- FUNCTION: Is required not only for elongation of protein synthesis
CC but also for the initiation of all mRNA translation through
CC initiator tRNA (fMet) aminoacylation (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + L-methionine + tRNA (Met) = AMP +
CC diphosphate + L-methionyl-tRNA (Met).
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC MetG subfamily 1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP000060; BAA8014.1; -.
CC PIR; B72714; B72714.
CC HSP; P00959; IMEA.
CC HAMAP; MF 00098; -.
CC InterPro; IPR002300; tRNA-synt_1a.
CC InterPro; IPR001412; tRNA-synt_1.
CC InterPro; IPR002304; tRNA-synt_met.
CC Pfam; PF00133; tRNA-synt_1; 1.
CC PRINTS; PR01041; TRNASYNTHMET.
CC TIGRfams; TIGR00398; tmetg; 1.
CC PROSITE; PS00178; AA TRNA LIGASE I; FALSE NEG.
CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC Metal-binding; Zinc; Complete proteome.
CC SITE 11
CC SITE 334 "HIGH" REGION.
CC SITE 337 "XMSK" REGION.
CC BINDING 337 ATP (BY SIMILARITY).
CC METAL 143 ZINC (BY SIMILARITY).
CC METAL 146 ZINC (BY SIMILARITY).
CC METAL 156 ZINC (BY SIMILARITY).
CC METAL 159 ZINC (BY SIMILARITY).
CC SEQUENCE 572 AA; 66480 MW; 6E21C1CFB666E4E CRC64;

Query Match 52.7%; Score 38.5; DB 1; Length 572;
Best Local Similarity 46.2%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 5; Gaps 1;

QY 2 WTN1W-WW 9
Db 467 WMTWGGNTYWW 479

Best Local Similarity 62.5%; Pred. No. 2e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 1 EWTN1W 8
Db 277 EWKS-WW 283

RESULT 39
SYM_AERPE STANDARD; PRT; 794 AA.
AC P35867;
DT 01-JUN-1994 (Rel. 29, Created);
DT 28-FEB-2003 (Rel. 41, Last sequence update);
DT 28-FEB-2003 (Rel. 41, Last annotation update);
DE Hypothetical protein Cgl0966.
GN Cgl0966.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RX "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RT Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-368 FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RX MEDLINE=92236417; PubMed=1667221;
RA Seep-Feldhaus A.H., Kainowski J., Puehler A.;
RT "Molecular analysis of the Corynebacterium glutamicum lysyl gene
RT involved in lysine uptake.";
RL Mol. Microbiol. 5:2995-3005(1991).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP005277; BAB98359.1; -.
CC PIR; S18758; CAA42857.1; -.
CC InterPro; IPR007425; DUF471.
CC InterPro; IPR007426; DUF472.
CC Pfam; PF04330; DUF471; 1.
CC Pfam; PF04331; DUF472; 1.
CC Hypothetical protein; Transmembrane; Complete proteome.
CC TRANSMEM 34 54 POTENTIAL.
CC TRANSMEM 67 87 POTENTIAL.
CC TRANSMEM 99 119 POTENTIAL.
CC TRANSMEM 132 152 POTENTIAL.
CC TRANSMEM 257 277 POTENTIAL.
CC TRANSMEM 283 303 POTENTIAL.
CC TRANSMEM 315 335 POTENTIAL.
CC TRANSMEM 353 373 POTENTIAL.
CC TRANSMEM 421 441 POTENTIAL.
CC SEQUENCE 794 AA; 87209 MW; 698E79C75BCC99A4 CRC64;

Query Match 52.7%; Score 38.5; DB 1; Length 794;
Best Local Similarity 46.2%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 5; Gaps 1;

QY 2 WTN1W-WW 9
Db 467 WMTWGGNTYWW 479
```

```
RESULT 40
PCS_RHILO
ID PCS_RHILO STANDARD; PRT; 263 AA.
AC Q98MN3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Phosphatidylcholine synthase (EC 2.7.8.24) (CDP-diglyceride-choline
DE O-phosphatidyltransferase) (PC synthase).
GN PCS OR ML0506.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MAFF303099;
RA Kureko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
CC -1- FUNCTION: Condenses choline with CDP-diglyceride to produce
CC phosphatidylcholine and CMP.
CC -1- CATALYTIC ACTIVITY: CDP-diacylglycerol + choline = CMP +
CC phosphatidylcholine.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: Belongs to the CDP-alcchol phosphatidyltransferase
CC class-I family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF002395; BAB48080.1; -.
CC InterPro; IPR000462; CDP-OH_P trans.
CC Pfam; PF01066; CDP-OH_P trans; 1.
CC Transferase; Phospholipid biosynthesis; Transmembrane;
CC Complete proteome.
CC TRANSMEM 28 50 POTENTIAL.
CC TRANSMEM 113 135 POTENTIAL.
CC TRANSMEM 147 164 POTENTIAL.
CC TRANSMEM 169 191 POTENTIAL.
CC TRANSMEM 204 226 POTENTIAL.
CC TRANSMEM 241 263 POTENTIAL.
CC SEQUENCE 268 AA; 29526 MW; E2C115F15F453A20 CRC64;
Query Match 52.1%; Score 38; DB 1; Length 268;
Best Local Similarity 57.1%; Pred. No. 1.1e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 WTNVWW 8
Db 54 WTAMFW 60
|| :||
|| :||
RESULT 41
MTP2_PROVU
ID MTP2_PROVU STANDARD; PRT; 336 AA.
AC P11409;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE Modification methylase PvuII (EC 2.1.1.113) (N-4 cytosine-specific
DE methyltransferase PvuII) (M.PvuII).
GN Proteus vulgaris
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Proteus.
OX NCBI_TaxID=585;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 13315;
RA MEDLINE=89296464; PubMed=2662138;
RT Tao T., Walter J., Brennan K.J., Cotterman M.M., Blumenthal R.M.;
RT "Sequence, internal homology and high-level expression of the gene
RT for a DNA-(cytosine N4)-methyltransferase, M.Pvu II."
RL Nucleic Acids Res. 17:4161-4175(1989).
RN [2]
RP SEQUENCE OF 1-80 FROM N.A. AND REVISION TO 44.
RX MEDLINE=95331563; PubMed=7607491;
RA Adams G.M., Blumenthal R.M.;
RT "Gene pvuIIM: a possible modulator of PvuII endonuclease subunit
RT association."
RL Gene 157:193-199(1995).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX MEDLINE=97351137; PubMed=9207015;
RA Gong W., O'Gara M., Blumenthal R.M., Cheng X.;
RT "Structure of pvu II DNA-(cytosine N4) methyltransferase, an example
RT of domain permutation and protein fold assignment."
RL Nucleic Acids Res. 25:2702-2715(1997).
CC -1- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC CAGCTC, CAUSES SPECIFIC METHYLATION ON C-4 ON BOTH STRANDS, AND
CC PROTECTS THE DNA FROM CLEAVAGE BY THE PvuII ENDONUCLEASE.
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA cytosine = S-
CC adenosyl-L-homocysteine + DNA N(4)-methylcytosine.
CC -1- SIMILARITY: Belongs to the N4-methyltransferase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X13778; CAA32026.1; -.
CC InterPro; IPR002295; D21N6 mtfrase.
CC InterPro; IPR001091; Met trans CN4.
CC InterPro; IPR002941; N6/N4 Mtase.
CC Pfam; PF01555; N6 N4 Mtase; 1.
CC PRINTS; PR00506; D21N6MTFRASE.
CC PRINTS; PS00508; S21N4MTFRASE.
CC PROSITE; PS00093; N4 MTASE; 1.
CC Transferase; Methyltransferase; Restriction system; Repeat;
CC 3D-structure.
CC REPEAT 11 113 1.
CC REPEAT 44 293 2.
CC CONFLICT 44 44 D -> E (IN REF. 1).
CC STRAND 23 24
CC STRAND 28 32
CC STRAND 35 38
CC HELIX 39 41
CC STRAND 47 52
CC HELIX 68 86
CC STRAND 87 97
CC STRAND 101 102
CC TURN 103 104
CC STRAND 105 108
CC HELIX 111 121
CC TURN 122 123
CC STRAND 126 134
CC TURN 139 140
```

```

FT TURN 143
FT HELIX 148
FT TURN 149
FT STRAND 164
FT HELIX 173
FT STRAND 222
FT HELIX 232
FT TURN 241
FT TURN 252
FT HELIX 262
FT TURN 266
FT STRAND 271
FT TURN 272
FT TURN 273
FT TURN 276
FT HELIX 278
FT TURN 286
FT STRAND 289
FT HELIX 297
FT HELIX 305
FT HELIX 314
FT TURN 326
FT STRAND 330
FT TURN 332
SQ SEQUENCE 336 AA; 38365 MW; 384371A667F86B91 CRC64;

Query Match 52.1%; Score 38; DB 1; Length 336;
Best Local Similarity 27.3%; Pred. No. 1.4e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 12; Gaps 1;

QY 1 EWTN-----IWMWAK 10
|||
Db 144 EWNKRIKRVKQAVNIVWFESK 165
|||

RESULT 42
CBL4 ARATH STANDARD; PRT; 431 AA.
AC Q9LFW3;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE COBRA-like protein 4 precursor.
GN CBL4 OR ATSG15630 OR F14P8.10.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016721; PubMed=111330714;
RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
RA Nakazaki N., Naruo K., Okumura S., Shimpō S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strommatt C.,
RA Wagner-McPherson C., Wollam A., Yokum M., Bell M., Dedhia N.,
RA Farnell J., Shan R., Rodriguez M., Hoon See L., Vil D., Baker J.,
RA Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.A.,
RA Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
RA Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
RA Entian K.-D., Terry N., Hartley N., Bent E., Johnson S.,
RA Langham S.-A., McColough B., Robben J., Grymonprez B., Zimmermann W.,
RA Ramsperger U., Wedler H., Balke K., Wedler E., Peters S.,
RA van Staveren M., Birke W., Mooijman P., Klein Lankhorst R.,
RA Weitzensegger T., Bothe G., Rose M., Hauf J., Berneiser S., Hempel S.,

```

```

RA Feldpausch M., Lamberth S., Villarroel R., Gielen J., Ardiles W.,
RA Bents O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,
RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Fransz P.F.;
RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
thaliana.";
RL Nature 408:823-826(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20363093; PubMed=10907847;
RA Asamizu E., Nakamura Y., Sato S., Tabata S.;
RT "A large scale analysis of cDNA in Arabidopsis thaliana: generation of
12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries.";
RL DNA Res. 7:175-180(2000).
RN [3]
RP TISSUE SPECIFICITY.
RX MEDLINE=22263984; PubMed=12376623;
RA Roudier F., Schindelman G., DeSalle R., Benfey P.N.;
RT "The COBRA family of putative GPI-anchored proteins in Arabidopsis. A
new fellowship in expansion.";
RL Plant Physiol. 130:538-548(2002).
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
(Potential).
CC -!- TISSUE SPECIFICITY: Expressed in roots, stems, leaves, flowers and
siliques.
CC -!- SIMILARITY: Belongs to the COBRA family.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to erroneous
gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ391114; CAC01762.1; ALT SEQ.
DR EMBL; AY567723; -; NOT_ANNOTATED_CDS.
DR PIR; TS1392; TS1392.
DR InterPro; IPR006918; Phytochel_synth.
DR Pfam; PF04833; Phytochel_synth; 1.
KW Membrane; GPI-anchor; Glycoprotein; Signal; Lipoprotein.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 414 COBRA-LIKE PROTEIN 4.
FT PROPEP 415 431 REMOVED IN MATURE FORM (POTENTIAL).
FT LIPID 414 414 GPI-anchor amidated asparagine
(Potential).
FT CARBOHYD 29 29 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 154 154 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 162 162 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 226 226 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 306 306 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 340 340 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 431 AA; 48515 MW; B75A23BAA2848A21 CRC64;

Query Match 52.1%; Score 38; DB 1; Length 431;
Best Local Similarity 66.7%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 WTNWVWAK 10
|||
Db 64 WTLGWTWAK 72

RESULT 43
CAIT_ECO57 STANDARD; PRT; 504 AA.
AC Q8XA10;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)

```



```
10-OCT-2003 (Rel. 42, Last annotation update)
DE L-carnitine/gamma-butyrobetaine antipporter.
DE CAIT OR Z0046 OR ECS0043.
OS Escherichia coli O157:H7.
CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.C., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.P., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.M., Lim A., Dimalanta E.T., Poramouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Chnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tohe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: Catalyzes the exchange of L-carnitine for gamma-
CC butyrobetaine and related betaines (By similarity).
CC -!- PATHWAY: Carnitine metabolism (conversion of carnitine to
CC gamma-butyrobetaine).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (By similarity).
CC -!- SIMILARITY: Belongs to the bcct transporter (TC 2.A.15) family.
CC CaIt subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB005180; AAG54343.1; -
CC PIR; C90634; C90634.
CC
CC HAMAP; MF_01049; -; 1.
CC InterPro; IPR000060; BCCT_transporter.
CC Pfam; PF02028; BCCT; 1.
CC ProDom; PD010111; BCCT_transporter; 1.
CC TIGRFAMs; TIGR00842; bcct; 1.
CC PROSITE; PS01303; BCCT; 1.
CC Transprot; Antipport; Transmembrane; Inner membrane; Complete proteome.
CC TRANSMEM 13 30 POTENTIAL.
CC TRANSMEM 50 72 POTENTIAL.
CC TRANSMEM 92 114 POTENTIAL.
CC TRANSMEM 141 163 POTENTIAL.
CC TRANSMEM 195 214 POTENTIAL.
CC TRANSMEM 229 251 POTENTIAL.
CC TRANSMEM 264 286 POTENTIAL.
CC TRANSMEM 312 334 POTENTIAL.
CC TRANSMEM 347 366 POTENTIAL.
CC TRANSMEM 410 432 POTENTIAL.
CC TRANSMEM 445 467 POTENTIAL.
CC TRANSMEM 472 494 POTENTIAL.
CC SEQUENCE 504 AA; 56581 MW; 13F0FEB6882E0941 CRC64;
SQ
Query Match 52.1%; Score 38; DB 1; Length 504;
Best Local Similarity 71.4%; Pred. No. 2.1e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 3 TINIWWA 9
DB 48 TINVWGA 54

RESULT 44
CAIT_ECOL6 STANDARD; PRT; 504 AA.
AC P59333;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE L-carnitine/gamma-butyrobetaine antipporter.
DE CAIT OR C0049.
GN Escherichia coli O6.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388334; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Blattner F.R.;
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli."
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
CC -!- FUNCTION: Catalyzes the exchange of L-carnitine for gamma-
CC butyrobetaine and related betaines (By similarity).
CC -!- PATHWAY: Carnitine metabolism (conversion of carnitine to
CC gamma-butyrobetaine).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (By similarity).
CC -!- SIMILARITY: Belongs to the bcct transporter (TC 2.A.15) family.
CC CaIt subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE016755; AAN78547.1; -
CC HAMAP; MF_01049; -; 1.
CC InterPro; IPR000060; BCCT_transporter.
CC Pfam; PF02028; BCCT; 1.
CC ProDom; PD010111; BCCT_transporter; 1.
CC TIGRFAMs; TIGR00842; bcct; 1.
CC PROSITE; PS01303; BCCT; 1.
CC Transprot; Antipport; Transmembrane; Inner membrane; Complete proteome.
CC TRANSMEM 13 30 POTENTIAL.
CC TRANSMEM 50 72 POTENTIAL.
CC TRANSMEM 92 114 POTENTIAL.
CC TRANSMEM 141 163 POTENTIAL.
CC TRANSMEM 195 214 POTENTIAL.
CC TRANSMEM 229 251 POTENTIAL.
CC TRANSMEM 264 286 POTENTIAL.
CC TRANSMEM 317 339 POTENTIAL.
CC TRANSMEM 410 432 POTENTIAL.
CC TRANSMEM 445 467 POTENTIAL.
CC TRANSMEM 472 494 POTENTIAL.
CC SEQUENCE 504 AA; 56573 MW; A8522606D303FDB2 CRC64;
SQ
Query Match 52.1%; Score 38; DB 1; Length 504;
Best Local Similarity 71.4%; Pred. No. 2.1e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```



```

DR InterPro: IPR000060; BCCT_transporter.
DR Pfam: PF02028; BCCT; 1.
DR ProDom: PD010111; BCCT_transporter; 1.
DR TIGRfams: TIGR00842; bcct; 1.
DR PROSITE: PS01303; BCCT; FALSE NEG.
KW Transport; Antipporter; Transmembrane; Inner membrane.
FT TRANSMEM 13 30 POTENTIAL.
FT TRANSMEM 50 72 POTENTIAL.
FT TRANSMEM 92 114 POTENTIAL.
FT TRANSMEM 144 163 POTENTIAL.
FT TRANSMEM 184 206 POTENTIAL.
FT TRANSMEM 229 251 POTENTIAL.
FT TRANSMEM 264 286 POTENTIAL.
FT TRANSMEM 312 334 POTENTIAL.
FT TRANSMEM 347 364 POTENTIAL.
FT TRANSMEM 410 432 POTENTIAL.
FT TRANSMEM 445 467 POTENTIAL.
FT TRANSMEM 472 494 POTENTIAL.
SQ SEQUENCE 504 AA; 56317 MW; EE4E302263257A43 CRC64;

Query Match 52.1%; Score 38; DB 1; Length 504;
Best Local Similarity 71.4%; Pred. No. 2.1e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 TNIWWWA 9
Db 48 TNVWGWA 54
||:||||
|:|:|

RESULT 47
CAIT_SHIFL STANDARD; PRT; 504 AA.
AC P59335;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE L-carnitine/gamma-butyrobetaine antiporter.
GN CAIT OR SF0037 OR S0039.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
CX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157."
RT Nucleic Acids Res. 30:4432-4441(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22530274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T."
RT Infect. Immun. 71:2775-2786(2003).
CC -1- FUNCTION: Catalyzes the exchange of L-carnitine for gamma-
CC butyrobetaine and related betaines (By similarity).
CC -1- PATHWAY: Carnitine metabolism (conversion of carnitine to
CC gamma-butyrobetaine).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (By similarity).
CC -1- SIMILARITY: Belongs to the bcct transporter (TC 2.A.15) family.
CC Cait subfamily.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE015041; AAN41703.1; -.
DR EMBL: AE016978; AAP15584.1; -.
DR HAMAP: MF_01049; -.
DR InterPro: IPR000060; BCCT_transporter.
DR Pfam: PF02028; BCCT; 1.
DR ProDom: PD010111; BCCT_transporter; 1.
DR TIGRfams: TIGR00842; bcct; 1.
DR PROSITE: PS01303; BCCT; 1.
KW Transport; Antipporter; Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 13 30 POTENTIAL.
FT TRANSMEM 50 72 POTENTIAL.
FT TRANSMEM 92 114 POTENTIAL.
FT TRANSMEM 141 163 POTENTIAL.
FT TRANSMEM 195 214 POTENTIAL.
FT TRANSMEM 229 251 POTENTIAL.
FT TRANSMEM 264 286 POTENTIAL.
FT TRANSMEM 312 334 POTENTIAL.
FT TRANSMEM 347 366 POTENTIAL.
FT TRANSMEM 410 432 POTENTIAL.
FT TRANSMEM 445 467 POTENTIAL.
FT TRANSMEM 472 494 POTENTIAL.
SQ SEQUENCE 504 AA; 56595 MW; 65B839C3E252F134 CRC64;

Query Match 52.1%; Score 38; DB 1; Length 504;
Best Local Similarity 71.4%; Pred. No. 2.1e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 TNIWWWA 9
Db 48 TNVWGWA 54
||:||||
|:|:|

RESULT 48
CAIT_SALTI STANDARD; PRT; 505 AA.
AC Q829LI;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE L-carnitine/gamma-butyrobetaine antiporter.
GN CAIT OR STY0084 OR T0075.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
CX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;

```

RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.P., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.,
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
and Ct18.",
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: Catalyzes the exchange of L-carnitine for gamma-
CC butyrobetaine and related betaines (By similarity).
CC -!- PATHWAY: Carnitine metabolism (conversion of carnitine to
CC gamma-butyrobetaine).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (By similarity).
CC -!- SIMILARITY: Belongs to the bcct transporter (TC 2.A.15) family.
CC Cait subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AL627265; CAD01228.1; -.
CC EMBL: AB016834; AAO67808.1; -.
CC HAMAP: MF 01049; -.
CC InterPro: IPR000069; BCCT_transporter.
CC Pfam: PF02028; BCCT; 1.
CC ProDom: PD010111; BCCT_transporter; 1.
CC TIGRFAMs: TIGR00842; bcct; 1.
CC PROSITE: PS01303; BCCT; 1.
KW Transport; Antiport; Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 13 30 POTENTIAL.
FT TRANSMEM 50 72 POTENTIAL.
FT TRANSMEM 92 114 POTENTIAL.
FT TRANSMEM 144 163 POTENTIAL.
FT TRANSMEM 195 214 POTENTIAL.
FT TRANSMEM 229 251 POTENTIAL.
FT TRANSMEM 264 286 POTENTIAL.
FT TRANSMEM 312 334 POTENTIAL.
FT TRANSMEM 347 366 POTENTIAL.
FT TRANSMEM 410 432 POTENTIAL.
FT TRANSMEM 445 467 POTENTIAL.
FT TRANSMEM 472 494 POTENTIAL.
SQ SEQUENCE 505 AA; 56747 MW; 43D41A070CBE1C8 CRC64;

Query Match 52.1%; Score 38; DB 1; Length 505;
Best Local Similarity 71.4%; Pred. No. 2.1e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 TINIWWA 9
DB 48 TINVWGA 54

RESULT 49
CAIT_SALT
ID CAIT_SALT STANDARD; PRT; 505 AA.
AC Q8ZXL1
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE L-carnitine/gamma-butyrobetaine antiporter.
GN CAIT OR STW0074
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,

RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan R., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.,
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.",
RL Nature 413:852-856(2001).
CC -!- FUNCTION: Catalyzes the exchange of L-carnitine for gamma-
CC butyrobetaine and related betaines (By similarity).
CC -!- PATHWAY: Carnitine metabolism (conversion of carnitine to
CC gamma-butyrobetaine).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (By similarity).
CC -!- SIMILARITY: Belongs to the bcct transporter (TC 2.A.15) family.
CC Cait subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

CC EMBL: AB008697; AAL19038.1; -.
CC StyGene: SG????; cait.
CC HAMAP: MF 01049; -.
CC InterPro: IPR000060; BCCT_transporter.
CC Pfam: PF02028; BCCT; 1.
CC ProDom: PD010111; BCCT_transporter; 1.
CC TIGRFAMs: TIGR00842; bcct; 1.
CC PROSITE: PS01303; BCCT; 1.

KW Transport; Antiport; Transmembrane; Inner membrane; Complete proteome.

FT TRANSMEM 13 30 POTENTIAL.
FT TRANSMEM 50 72 POTENTIAL.
FT TRANSMEM 92 114 POTENTIAL.
FT TRANSMEM 144 163 POTENTIAL.
FT TRANSMEM 195 214 POTENTIAL.
FT TRANSMEM 229 251 POTENTIAL.
FT TRANSMEM 264 286 POTENTIAL.
FT TRANSMEM 312 334 POTENTIAL.
FT TRANSMEM 347 366 POTENTIAL.
FT TRANSMEM 410 432 POTENTIAL.
FT TRANSMEM 445 467 POTENTIAL.
FT TRANSMEM 472 494 POTENTIAL.
SQ SEQUENCE 505 AA; 56643 MW; 0E678698707F9665 CRC64;

Query Match 52.1%; Score 38; DB 1; Length 505;
Best Local Similarity 71.4%; Pred. No. 2.1e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 TINIWWA 9
DB 48 TINVWGA 54

RESULT 50
Y096_MYCG
ID Y096_MYCG STANDARD; PRT; 650 AA.
AC P47342; Q49188;
DT 01-FEB-1996 (Rel. 33, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG096.
GN MG096.

OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,

RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RP REVISIONS.
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischman R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.,
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP PARTIAL SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; Pubmed=8253683;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
RN [2]
CC -1- SIMILARITY: BELONGS TO THE MG032 / MG096 / MG288 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U99690; AAC71314.1; -;
DR EMBL; U01713; AAC43186.1; -;
DR EMBL; U01762; AAD10577.1; -;
DR TIGR; MG096; -;
DR InterPro; IPR004306; MG032/096/288_1.
DR Pfam; PF03072; DUF237; 1.
DR Pfam; PF03086; DUF240; 1.
DR ProDom; PDC04834; MG032/096/288_2; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 650 AA; 74802 MW; 30539DDC6F69DDC3 CRC64;

Query Match 52.1%; Score 38; DB 1; Length 650;
Best Local Similarity 57.1%; Pred. No. 2.7e+02;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 WTNIWW 8
DB 538 WYVHW 544

RESULT 51
OPGH CAUCR
ID OPGH CAUCR STANDARD; PRT; 663 AA.
AC Q9A6R5;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glucans biosynthesis glucosyltransferase H (EC 2.4.1.-).
GN OPGH OR CC2018
OS Caulobacter crescentus.
CC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
CC Caulobacteraceae; Caulobacter.
CX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; Pubmed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,

RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Winn M.L., Haft D.H.,
RA Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of *Caulobacter crescentus*.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -1- FUNCTION: Involved in the biosynthesis of osmoregulated
CC periplasmic glucans (OPGs) (By similarity).
CC -1- PATHWAY: OPG (osmoregulated periplasmic glucans) biosynthesis.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (By similarity).
CC -1- SIMILARITY: Belongs to the glucosyltransferase family 2. OpgH
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE005875; AAK23993.1; -;
DR PIR; E87499; E87499.
DR TIGR; CC2018; -;
DR HAMAP; MF_01072; -; 1.
DR InterPro; IPR001173; Glyco_trans_2.
DR Pfam; PF00535; Glycosyltransf_2; 1.
KW Transferase; Glycosyltransferase; Inner membrane; Transmembrane;
KW Complete proteome.
FT TRANSMEM 64 86 Potential.
FT TRANSMEM 101 123 Potential.
FT TRANSMEM 413 435 Potential.
FT TRANSMEM 470 492 Potential.
FT TRANSMEM 558 580 Potential.
FT TRANSMEM 584 606 Potential.
SQ SEQUENCE 663 AA; 72442 MW; A57C91277B09CD96 CRC64;

Query Match 52.1%; Score 38; DB 1; Length 663;
Best Local Similarity 50.0%; Pred. No. 2.7e+02;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 WTNIWW 9
DB 300 WTGLAWS 307

RESULT 52
BETT_ECOLI
ID BETT_ECOLI STANDARD; PRT; 677 AA.
AC P17447;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE High-affinity choline transport protein.
GN BETT OR B0314 OR Z0401 OR ECS0360.
OS *Escherichia coli*, and
OS *Escherichia coli* O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; *Escherichia*.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92065800; Pubmed=1956285;
RA Lamark T., Kaasen E., Eshoo M.W., Falkenberg P., McDougall J.,
RA Strom A.R.;
RT "DNA sequence and analysis of the bet genes encoding the
RT osmoregulatory choline-glycine betaine pathway of *Escherichia coli*.";
RL Mol. Microbiol. 5:1049-1064(1991).
RN [2]
RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;
RX MEDLINE=37426517; PubMed=9278503;
RA Sittner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RA Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
RA Federspiel N., Hyman K., Kalman S., Komp C., Kurdi O., Lew H.,
RA Lin D., Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharam T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 403:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RA STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Lida T., Takami H., Honda T., Sasaki C., Ogawara N., Yasunaga T.,
RA Kunara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12";
RL DNA Res. 8:11-22(2001).
CC -!- FUNCTION: High-affinity uptake of choline driven by a proton-
CC motive force.
CC -!- PATHWAY: Osmoregulatory choline-glycine betaine pathway.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC -!- INDUCTION: BY osmotic stress. Choline is required for full
CC expression.
CC -!- SIMILARITY: Belongs to the bccT transporter (TC 2.A.15) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X52905; CAA37090.1; -;
CC EMBL; AE000138; AAC73417.1; -;
CC EMBL; U73857; AAB18040.1; -;
CC EMBL; AE005209; AAG54656.1; -;
CC EMBL; AP002551; BAB33783.1; -;
CC PIR; D85524; D85524.
CC PIR; H90673; H90673.
CC PIR; S15179; S15179.
CC EcoGene; EG10112; betT.
CC InterPro; IPR000060; BCC_Transporter.
CC Pfam; PF02228; BCC_T; 1.
CC ProDom; PDC10111; BCC_Transporter; 1.
CC TIGRFAMs; TIGR00842; bccT; 1.
CC PROSITE; PS01303; BCC_T; 1.
CC Transport; Transmembrane; Inner membrane; Complete proteome.
CC TRANSMEM 15 35 POTENTIAL.
CC TRANSMEM 54 74 POTENTIAL.
CC TRANSMEM 94 114 POTENTIAL.
CC TRANSMEM 144 164 POTENTIAL.
CC TRANSMEM 196 216 POTENTIAL.

FT TRANSMEM 233 253 POTENTIAL.
FT TRANSMEM 265 285 POTENTIAL.
FT TRANSMEM 319 339 POTENTIAL.
FT TRANSMEM 350 370 POTENTIAL.
FT TRANSMEM 412 432 POTENTIAL.
FT TRANSMEM 452 472 POTENTIAL.
FT TRANSMEM 477 497 POTENTIAL.
SQ SEQUENCE 677 AA; 75842 MW; 76062FCB877D0A55 CRC64;

Query Match 52.1%; Score 38; DB 1; Length 677;
Best Local Similarity 42.9%; Pred No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 6; Gaps 1;

QY 1 EWTNIW-----WW 8
||| |||
DB 314 EWMNNWTLFFWAWW 327

RESULT 53
TREY MYCTU
ID TREY MYCTU STANDARD; PRT: 765 AA.
AC Q10768;
DT 01-OCT-1996 (rel. 34, Created)
DT 01-OCT-1996 (rel. 34, Last sequence update)
DT 10-OCT-2003 (rel. 42, Last annotation update)
DE Potential maltotriose trehalose synthase [EC 5.4.99.15] (1,4)-
DE alpha-D-glucan 1-alpha-D-glucosylmutase.
DE RVI563C OR MT1614 OR MTCY48.02.
GN Mycobacterium tuberculosis.
OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies K., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.S., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- FUNCTION: Catalyzes the conversion of maltotriose into
CC the non-reducing saccharide, maltotriose trehalose (alpha-
CC maltotriose) alpha-D-glucoside) by intramolecular
CC transglycosylation (By similarity).
CC -!- CATALYTIC ACTIVITY: 4-((1,4)-alpha-D-glucosyl)(N-1)-D-glucose = 1-
CC alpha-D-((1,4)-alpha-D-glucosyl)(N-1)-alpha-D-glucopyranoside.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SIMILARITY: Belongs to family 13 of glycosyl hydrolases.
CC -!- SIMILARITY: TO M.TUBERCULOSIS RVI562C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; Z74020; CAA98328.1; -;
 DR EMBL; AE007027; AAK45881.1; -;
 DR PIR; H70763; H70763.
 DR TIGR; M1614; -;
 DR Tuberculist; Rv1563c; -;
 DR InterPro; IPR006347; Alpha_amyl_cat.
 DR Pfam; PF00128; alpha-amylase; 1.
 KW Hypothetical protein; Isomerase; Complete proteome.
 SQ SEQUENCE 765 AA; 82706 MW; 34E9A5CB5624DeA CRC64;

Query Match 52.1%; Score 38; DB 1; Length 765;
 Best Local Similarity 80.0%; Pred. No. 3.1e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 NIWMM 8 PRT; 856 AA.

Db 107 NAWMM 111

RESULT 54

ENV_HV1PV ID ENV HV1PV STANDARD; PRT; 856 AA.

AC P03376;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Envelope polyprotein GP160 precursor [Contains: Exterior membrane
 DE glycoprotein (GP120); Transmembrane glycoprotein (GP41)].
 GN ENV.

OS Human immunodeficiency virus type 1 (PV22 isolate) (HIV-1).
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11700;

RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=85111157; PubMed=2982104;
 RA Muesing M.A., Smith D.H., Cabradilla C.D., Berton C.V., Lasky L.A.,
 RA Capon D.J.;
 RT "Nucleic acid structure and expression of the human
 RT AIDS/lymphadenopathy retrovirus.";
 RL Nature 313:450-458 (1985).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; K02083; AAB59873.1; -;
 DR EMBL; X01762; CAA25903.1; ALT_SEQ.
 DR PIR; A03974; VCLJVL.
 DR HIV; K02083; ENVSPV22.

DR InterPro; IPR000328; Env GP41.

DR Pfam; PF00516; GP120; 1.

DR Pfam; PF00517; GP41; 1.

KW AIDS; Coat protein; Polyprotein; Glycoprotein; Transmembrane;
 KW Signal.

FT CHAIN 1 30
 FT CHAIN 31 511
 FT CHAIN 512 856

FT DISULFID 54 74
 FT DISULFID 119 205
 FT DISULFID 126 196
 FT DISULFID 131 157
 FT DISULFID 218 247
 FT DISULFID 228 239
 FT DISULFID 296 331

EXTERIOR MEMBRANE GLYCOPROTEIN.
 TRANSMEMBRANE GLYCOPROTEIN.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.

FT DISULFID 378 445
 FT DISULFID 385 418
 FT CARBOHYD 88
 FT CARBOHYD 136
 FT CARBOHYD 141
 FT CARBOHYD 156
 FT CARBOHYD 160
 FT CARBOHYD 186
 FT CARBOHYD 197
 FT CARBOHYD 230
 FT CARBOHYD 234
 FT CARBOHYD 241
 FT CARBOHYD 262
 FT CARBOHYD 276
 FT CARBOHYD 289
 FT CARBOHYD 295
 FT CARBOHYD 301
 FT CARBOHYD 332
 FT CARBOHYD 339
 FT CARBOHYD 356
 FT CARBOHYD 386
 FT CARBOHYD 392
 FT CARBOHYD 397
 FT CARBOHYD 406
 FT CARBOHYD 448
 FT CARBOHYD 463
 FT CARBOHYD 611
 FT CARBOHYD 616
 FT CARBOHYD 625
 FT CARBOHYD 637
 FT CARBOHYD 674
 FT CARBOHYD 750
 FT CARBOHYD 816
 SQ SEQUENCE 856 AA; 97339 MW; 5FCDBDC3C1209B3 CRC64;

Query Match 52.1%; Score 38; DB 1; Length 856;
 Best Local Similarity 50.0%; Pred. No. 3.5e+02;
 Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNIWMM 8

Db 665 KWANLWMM 672

RESULT 55

ENV_HV1ZH ID ENV HV1ZH STANDARD; PRT; 856 AA.

AC P05881;
 DT 01-NOV-1988 (Rel. 09, Created)

DT 01-NOV-1988 (Rel. 09, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Envelope polyprotein GP160 precursor [Contains: Exterior membrane
 DE glycoprotein (GP120); Transmembrane glycoprotein (GP41)].
 GN ENV.

OS Human immunodeficiency virus type 1 (Zaire H2321 isolate) (HIV-1).
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11692;
 RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=89228766; PubMed=2713163;
 RA Srinivasan A., York D., Butler D. Jr., Jannoun-Nasr R., Getchell J.,
 RA McCormack J., Ou C.Y., Myers G., Smith T., Chen E.;
 RT "Molecular characterization of HIV-1 isolated from a serum collected
 RT in 1976: nucleotide sequence comparison to recent isolates and
 RT generation of hybrid HIV.";
 RL AIDS Res. Hum. Retroviruses 5:121-129 (1989).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

SQ SEQUENCE 865 AA; 97809 MW; 28828BC2314ADCAC CRC64;

Query Match 52.1%; Score 38; DB 1; Length 865;
 Best Local Similarity 50.0%; Pred. No. 3.5e+02;
 Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNINWW 8
 DB 674 KWANLWNW 681

RESULT 57

HGBC HAEIN STANDARD; PRT; 993 AA.

AC Q9KIV0;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemoglobin binding protein C precursor.
 GN HGBC.
 OS Haemophilus influenzae.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 CC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NTHI N182;
 RX MEDLINE=20316037; PubMed=10858226;
 RA Cope L.D., Hrkal Z., Hansen E.J.;
 RT "Detection of phase variation in expression of proteins involved in
 hemoglobin and hemoglobin-haptoglobin binding by nontypeable
 Haemophilus influenzae.";
 RL Infect. Immun. 68:4092-4101 (2000).
 CC -!- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OF THE HUMAN HOST AND
 IS REQUIRED FOR HEME UPTAKE.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
 EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCAA
 REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAIRING.
 CC ADDITION OR LOSS OF CCAA REPEAT UNITS WOULD CHANGE THE READING
 FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
 REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A
 WAY TO AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST.
 CC -!- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
 HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).

EMBL; AF221060; AAF80178.1;
 InterPro: IPR006970; PT.
 InterPro: IPR000531; TonB_boxC.
 Pfam; PF04886; PT; 1.
 Pfam; PF00593; TonB_dep_Rec; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Outer membrane; Transport; TonB_box; Multigene family; Signal;
 Receptor; Repeat.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 993 HEMOGLOBIN BINDING PROTEIN C.
 FT DOMAIN 26 37 3 X 4 AA TANDEM REPEATS OF Q-P-T-N.
 FT REPEAT 26 29 1.
 FT REPEAT 30 33 2.
 FT REPEAT 34 37 3.
 FT REPEAT 38 41 3.
 FT REPEAT 42 45 5.
 FT REPEAT 46 49 6.
 FT REPEAT 50 53 7.
 FT REPEAT 54 57 8.
 FT SITE 976 993 TONB C-TERMINAL BOX.
 SQ SEQUENCE 993 AA; 113616 MW; A5513F3B2C64:612 CRC64;

Query Match 52.1%; Score 38; DB 1; Length 993;
 Best Local Similarity 62.5%; Pred. No. 4.1e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 TNLWVWAK 10
 DB 562 TDVQWVAK 569

RESULT 58

HGBA HAEIN STANDARD; PRT; 1013 AA.

AC Q9KIV2;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemoglobin binding protein A precursor.
 GN HGBA.
 OS Haemophilus influenzae.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 CC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NTHI N182;
 RX MEDLINE=20316037; PubMed=10858226;
 RA Cope L.D., Hrkal Z., Hansen E.J.;
 RT "Detection of phase variation in expression of proteins involved in
 hemoglobin and hemoglobin-haptoglobin binding by nontypeable
 Haemophilus influenzae.";
 RL Infect. Immun. 68:4092-4101 (2000).
 CC -!- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OF THE HUMAN HOST AND
 IS REQUIRED FOR HEME UPTAKE.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
 EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCAA
 REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAIRING.
 CC ADDITION OR LOSS OF CCAA REPEAT UNITS WOULD CHANGE THE READING
 FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
 REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A
 WAY TO AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST.
 CC -!- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
 HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).

EMBL; AF221059; AAF80176.1;
 InterPro: IPR006970; PT.
 InterPro: IPR000531; TonB_boxC.
 Pfam; PF04886; PT; 1.
 Pfam; PF00593; TonB_dep_Rec; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Outer membrane; Transport; TonB_box; Multigene family; Signal;
 Receptor; Repeat.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 1013 HEMOGLOBIN BINDING PROTEIN A.
 FT DOMAIN 26 57 8 X 4 AA TANDEM REPEATS OF Q-P-T-N.
 FT REPEAT 26 29 1.
 FT REPEAT 30 33 2.
 FT REPEAT 34 37 3.
 FT REPEAT 38 41 4.
 FT REPEAT 42 45 5.
 FT REPEAT 46 49 6.
 FT REPEAT 50 53 7.
 FT REPEAT 54 57 8.
 FT SITE 567 574 TONB BOX.

FT SIZE 996 1013 TONE C-TERMINAL BOX.
 SQ SEQUENCE 1013 AA; 116260 MW; 769964335A4ED3C1 CRC64;

Query Match 52.1%; Score 38; DB 1; Length 1013;
 Best Local Similarity 62.5%; Pred. No. 4.1e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 TNIWWAK 10
 Db 582 TDVQWAK 589

RESULT 59
 ID PD22 ARATH STANDARD; PRT; 1039 AA.
 AC Q3M9W8;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE Phospholipase D p2 (EC 3.1.4.4) (AtPLDp2) (Phospholipase D2 PHOX and
 DE PK containing domain) (Phospholipase D zeta 2) (PLDzeta2).
 GN PLD2 OR AT3G05630 OR F18C1.10.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016720; PubMed=11130713;
 RA Salanoubat M., Lemcke K., Rieger M., Ansoorge W., Unseld M.,
 RA Farwamm B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
 RA Delsen M., Bottry M., Grivell L.A., Mache R., Puigdomenech P.,
 RA De Simone V., Choise N., Attiguenave F., Robert C., Brottier P.,
 RA Wincker P., Cattolico J., Weissenbach J., Saurin W., Quetier F.,
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
 RA Wurnbach E., Drzonek H., Erflie H., Jordan N., Bangert S.,
 RA Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
 RA Vezi A., D'Angelo M., Pallavicini A., Toppi S., Simonati B.,
 RA Conrad A., Hornischer K., Kauer G., Loebner T.-H., Nordsiek G.,
 RA Reichelt J., Scharfe M., Schoen O., Bargues M., Terol J., Climent J.,
 RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
 RA Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masny D.,
 RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
 RA Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,
 RA Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P.,
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.O., Jenkins J.,
 RA Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
 RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
 RA Pal G., Millscher J., Sellers P., Gill J.E., Feldblyum T.V.,
 RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
 RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
 RA Sasamoto S., Kimura T., Igesawa K., Kawashima K., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakayama S., Nakazaki N., Shino S., Takeuchi C., Wada T.,
 RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
 thaliana".
 RL Nature 408:920-922(2000).
 CC -!- FUNCTION: Hydrolyzes glycerol-phospholipids at the terminal
 CC phosphodiesteric bond. Phosphatidylcholine-selective.
 CC -!- CATALYTIC ACTIVITY: A phosphatidylcholine + a
 CC phosphate.
 CC -!- ENZYME REGULATION: Calcium-independent and PIP2-dependent.
 CC -!- SIMILARITY: Belongs to the phospholipase D family. XPX-PLD
 CC subfamily.
 CC -!- SIMILARITY: Contains 1 phox homology (PX) domain.
 CC -!- SIMILARITY: Contains 1 PH domain.
 CC -!- SIMILARITY: Contains 2 PH phosphodiesterase domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 or send an email to license@isb-sib.ch).

 DR EMBL; AC011620; AAF26134.1; -;
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR001736; PLD.
 DR InterPro; IPR001683; PX.
 DR Pfam; PF00169; PH; 1.
 DR Pfam; PF00614; PLDC; 2.
 DR SMART; SM00233; PH; 1.
 DR SMART; SM00155; PLDC; 2.
 DR PROSITE; PSS0003; PH DOMAIN; 1.
 DR PROSITE; PSS0035; PLD; 2.
 DR PROSITE; PSS0195; PX; FALSE NEG.
 KW Hydrolase; Lipid degradation; Repeat; Multigene family.
 FT DOMAIN 45 205 PX.
 FT DOMAIN 215 343 PH.
 FT DOMAIN 465 492 PLD PHOSPHODIESTERASE 1.
 FT DOMAIN 840 867 PLD PHOSPHODIESTERASE 2.
 FT ACT_SITE 470 470 POTENTIAL.
 FT ACT_SITE 472 472 POTENTIAL.
 FT ACT_SITE 477 477 POTENTIAL.
 FT ACT_SITE 845 845 POTENTIAL.
 FT ACT_SITE 847 847 POTENTIAL.
 FT ACT_SITE 852 852 POTENTIAL.
 SQ SEQUENCE 1039 AA; 117902 MW; A8433C237C3B77E3 CRC64;

Query Match 52.1%; Score 38; DB 1; Length 1039;
 Best Local Similarity 62.5%; Pred. No. 4.2e+02;
 Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 EWTNIWW 8
 Db 671 EAVNDWW 678

RESULT 60
 ID POLG HRV89 STANDARD; PRT; 2164 AA.
 AC Q87210; Q82096; Q82097; Q82098; Q82099; Q82100; Q82101; Q82102;
 AC Q82103; Q82104; Q82105;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Genome polyprotein [Contains: Coat protein VP4 (P1A); Coat protein VP2
 DE (P1B); Coat protein VP3 (P1C); Coat protein VP1 (P1D); Core protein
 DE P2A; Core protein P2B; Core protein P2C; Core protein P3A; Genome-
 DE linked protein VP3 (P3B); Picornain 3C (EC 3.4.22.28) (Protease 3C)
 DE (P3C); RNA-directed RNA polymerase P3D (EC 2.7.7.48)].
 OS Human rhinovirus 89 (HRV-89).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
 OC Rhinovirus.
 OC NCBI_TaxID=12132;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87204179; PubMed=3033653;
 RA Duechler M., Skern T., Sommergruber W., Neubauer C., Gruendler P.,
 RA Fogy I., Blas D., Kuechler E.;
 RT "Evolutionary relationships within the human rhinovirus genus:
 RT comparison of serotypes 89, 2, and 14".
 RL Proc. Natl. Acad. Sci. U.S.A. 84:2605-2609(1987).
 CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
 CC Q/G SITES IN THE POLYPEPTIDE. IT MAY BE A CYSTEINE PROTEASE.
 CC -!- CATALYTIC ACTIVITY: Selective cleavage of Gln-Gly bond in the
 CC poliovirus polyprotein. In other picornavirus reactions Glu may be
 CC substituted for Gln, and Ser or Thr for Gly.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA) (N).
 CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units,
 CC each of which is composed of one copy each of proteins VP1, VP2,

CC VP3, and VP4.
 CC -1- P1M: Specific enzymatic cleavages in vivo yield mature proteins.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M16248; AAA5762.1; -
 CC EMBL: AL0937; CAA00931.1; -
 CC MEROPS: C03.007; -
 CC MEROPS: C03.021; -
 CC InterPro: IPR004004; Calici_pol_hel.
 CC InterPro: IPR009003; Cys_Ser_trypsin.
 CC InterPro: IPR000199; Pept_3C_picorn.
 CC InterPro: IPR000081; Peptidase_C3.
 CC InterPro: IPR003138; Pico_P1A.
 CC InterPro: IPR002527; Pico_P2B.
 CC InterPro: IPR001676; Rhv_P2B.
 CC InterPro: IPR000655; RNA_helicase.
 CC InterPro: IPR007095; RNA_pol_DS_PS.
 CC InterPro: IPR001205; RNA_pol_P3D.
 CC InterPro: IPR007094; RNA_pol_P5Vir.
 CC InterPro: IPR008975; Viral_cap_coat.
 CC Pfam: PF00548; Cys-protease_3C; 1.
 CC Pfam: PF02226; Pico_P1A; 1.
 CC Pfam: PF00947; Pico_P2A; 1.
 CC Pfam: PF01552; Pico_P2B; 1.
 CC Pfam: PF00073; rhv; 3.
 CC Pfam: PF00680; RNA_dep_RNA_pol; 1.
 CC Pfam: PF03910; RNA_helicase; 1.
 CC PRINTS: PR00918; CALICIVIRUSNS.
 CC ProDom: PD001125; Cys_protease_3C; 1.
 CC ProDom: PD001306; Pico_P2A; 1.
 CC ProDom: PD001274; Pico_P2B; 1.
 CC Polyprotein: Coat protein; Core protein; Transferase;
 CC RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate;
 CC Lipoprotein.
 CC CHAIN 1 69 COAT PROTEIN VP4.
 CC CHAIN 70 336 COAT PROTEIN VP2.
 CC CHAIN 337 574 COAT PROTEIN VP3.
 CC CHAIN 575 866 COAT PROTEIN VP1.
 CC CHAIN 867 1008 CORE PROTEIN P2A.
 CC CHAIN 1009 1103 CORE PROTEIN P2B.
 CC CHAIN 1104 1424 CORE PROTEIN P2C.
 CC CHAIN 1425 1500 CORE PROTEIN P3A.
 CC CHAIN 1501 1521 GENOME-LINKED PROTEIN VP6.
 CC CHAIN 1522 1704 PICORNAIN 3C.
 CC CHAIN 1705 2164 RNA-DIRECTED RNA POLYMERASE P3D.
 CC LIPID 2 2 N-myristoyl glycine (in host) (By
 CC similarity).
 CC ACT_SITE 1668 1668 PROTEASE (POTENTIAL).
 CC ACT_SITE 1682 1682 PROTEASE (POTENTIAL).
 CC SEQUENCE 2164 AA; 241063 MW; F5D9C8F4FBEA7D54 CRC64;
 CC
 CC Query Match 51.4%; Score 37.5; DB 1; Length 2164;
 CC Best Local Similarity 50.0%; Pred. No. 1e+03;
 CC Matches 5; Conservative 1; Mismatches 1; Indels 3; Gaps 1;
 CC
 CC QY 2 WTNIN---WWW 8
 CC ||: |||
 CC Db 140 WTSASSGWW 149
 CC
 CC RESULT 61
 CC LPW_VIBPA
 CC ID LPW_VIBPA STANDARD; PRT; 41 AA.
 CC AC P22100;
 CC DT 01-AUG-1991 (Rel. 19, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Tri operon leader peptide.
 GN TRPL OR VP1955.
 OS Vibrio parahaemolyticus.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=670;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BB22;
 RA MEDLINE=92127059; PubMed=1773058;
 RX Crawford I.P., Han C.Y., Silverman M.;
 RT "Sequence and features of the tryptophan operon of Vibrio
 RT parahaemolyticus.";
 RL DNA Seq. 1:189-196(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIMD 2210633 / Serotype O3:K6;
 RX MEDLINE=22508454; PubMed=12620739;
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
 RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 RA Yasunaga T., Honda T., Shinagawa H., Battori M., Iida T.;
 RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
 RT distinct from that of V. cholerae.";
 RL Lancet 361:743-749(2003).
 CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN CONTROL OF THE BIOSYNTHESIS
 CC OF TRYPTOPHAN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X17149; CAA35030.1; -
 CC EMBL: AP005079; BAC60218.1; -
 CC KW Tryptophan biosynthesis; Leader peptide; Complete proteome.
 FT CONFLICT 23 23 T->A (IN REF. 1).
 SQ SEQUENCE 41 AA; 4968 MW; DFE9AB32F3869349 CRC64;
 CC
 CC Query Match 50.7%; Score 37; DB 1; Length 41;
 CC Best Local Similarity 66.7%; Pred. No. 25;
 CC Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC QY 2 WTNLWW 7
 CC ||: |||
 CC Db 31 WTSWW 36
 CC
 CC RESULT 62
 CC RL28_Syny3
 CC ID RL28_Syny3 STANDARD; PRT; 78 AA.
 AC P72851;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 50S ribosomal protein L28.
 GN RPL28 OR RPL28 OR SSR1604.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=97061201; PubMed=8905231;
 RX Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
 RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
 RA Yamada M., Yaeuda M., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium

RT Strychnocystis sp. strain PCC6803. II. Sequence determination of the
entire genome and assignment of potential protein-coding regions.*;
DNA Res. 3:109-136(1996).
CC -!- SIMILARITY: Belongs to the L28p family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; D90301; BAA16866.1; -
CC PIR; S74715; S74715.
CC HAWAP; MF_C0373; -; 1.
CC InterPro; IPR001383; Ribosomal L28.
CC Pfam; PF00830; Ribosomal L28; 1.
CC TIGRFAMs; TIGR00009; L28; 1.
CC Ribosomal protein; Complete proteome.
CC SEQUENCE 78 AA; 8993 MW; 18EC460566C80199 CRC64;
CC
CC Query Match 50.7%; Score 37; DB 1; Length 78;
Best Local Similarity 42.9%; Pred. No. 46;
Matches 3; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
CC
CC QY 1 EWTNIWV 7
DB 34 QWKRVMW 40
CC
CC RESULT 63
CC YRH1 RHOSH STANDARD; PRT; 165 AA.
CC ID YRH1 RHOSH
CC AC Q53229; Q53230; Q53231;
CC DT 30-MAY-2000 (Rel. 39, Created)
CC DT 30-MAY-2000 (Rel. 39, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE Hypothetical 19.2 kDa protein in RHO 5' region (ORF1).
CC OS Rhodobacter sphaeroides (Rhodospirillum rubrum).
CC OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
CC CC Rhodobacteraceae; Rhodobacter.
CC NCBI_TaxID=1063;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
CC MEDLINE=96178958; PubMed=8606169;
CC RA Gomelsky M., Kaplan S.;
CC RT "The Rhodobacter sphaeroides 2.4.1 rho gene: expression and genetic
analysis of structure and function."
CC RL J. Bacteriol. 178:1946-1954(1996).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0093 FAMILY.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-3 OR MET-12 IS THE
INITIATOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; L76097; AAB02031.1; -
CC EMBL; L76097; AAB02032.1; ALT INIT.
CC InterPro; IPR005265; Cons_hypoth701.
CC Pfam; PF03653; UPF0093; 1.
CC TIGRFAMs; TIGR00701; TIGR00701; 1.
CC Hypothetical protein; Transmembrane.
CC TRANSMEM 26 46
CC TRANSMEM 77 97
CC POTENTIAL.

FT TRANSMEM 145 165 POTENTIAL.
SQ SEQUENCE 165 AA; 19136 MW; 2474FCFDC2CC8 CRC64;
CC
CC Query Match 50.7%; Score 37; DB 1; Length 165;
Best Local Similarity 40.0%; Pred. No. 96;
Matches 4; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
CC
CC QY 1 EWTNIWVWAK 10
DB 96 DMSMLPWTWK 105
CC
CC RESULT 64
CC SUR4 CAEEL STANDARD; PRT; 277 AA.
CC ID SUR4 CAEEL
CC AC Q18864; O18006;
CC DT 01-NOV-1997 (Rel. 35, Created)
CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Surfeit locus protein 4 homolog.
CC GN SFT-4 OR SURF-4 OR C54H2.5.
CC OS Caenorhabditis elegans.
CC CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
CC OC Rhabditidae; Peloderinae; Caenorhabditis.
CC NCBI_TaxID=6239;
CC [1]
CC SEQUENCE FROM N.A.
CC MEDLINE=96413309; PubMed=8816471;
CC RA Ames N., Fried M.;
CC RT "Surfeit locus gene homologs are widely distributed in invertebrate
genomes."
CC RL Mol. Cell. Biol. 16:5591-5596(1996).
CC [2]
CC SEQUENCE FROM N.A.
CC STRAIN=Bristol N2;
CC RA Fulton L., Gattung S.;
CC RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
reticulum (by similarity).
CC -!- SIMILARITY: Belongs to the SURF4 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; U58728; AAB00591.1; -
CC EMBL; Y14949; CAA75173.1; -
CC FIR; T29611; T29611.
CC WormPep; C54H2.5; CE06987.
CC InterPro; IPR002995; Surf4.
CC Pfam; PF02077; SURF4; 1.
CC ProDom; PD010195; Surf4; 1.
CC PROSITE; PS01339; SURF4; 1.
CC Transmembrane; Endoplasmic reticulum.
CC TRANSMEM 71 91
CC TRANSMEM 97 117
CC TRANSMEM 118 138
CC TRANSMEM 163 183
CC TRANSMEM 187 207
CC TRANSMEM 213 233
CC TRANSMEM 250 270
CC SITE 274 275
CC ENDOPLASMIC RETICULUM RETRIEVAL MOTIF
{POTENTIAL}
CC CONFLICT 17 17 A -> R (IN REF. 1).
CC SEQUENCE 277 AA; 31813 MW; 8FF2848676CC0245 CRC64;
CC
CC Query Match 50.7%; Score 37; DB 1; Length 277;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 WTNTNW 7
 DB 226 WLNWW 231

RESULT 65
 IASP HUMAN
 ID IASP HUMAN STANDARD; PRT; 407 AA.
 AC Q8WU5; Q8Y230;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE RelA-associated inhibitor (Inhibitor of ASPP protein) (Protein iASPP)
 DE (PPI13B-like protein).
 GN RAI OR IASPP OR PPI13BL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 [1]
 RN SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, TISSUE
 RP SPECIFICITY, AND INTERACTION WITH RELA.
 RC TISSUE=Placenta;
 RX MEDLINE=99269105; PubMed=10336463;
 RA Yang J.-P., Hori M., Sanda T., Okamoto T.;
 RT Identification of a novel inhibitor of nuclear factor-kappaB, RelA-
 RT associated inhibitor.;
 RL J. Biol. Chem. 274:15662-15670 (1999).
 [2]
 RN SEQUENCE OF 141-407 FROM N.A.
 RP TISSUE=Liver;
 RC MEDLINE=22388257; PubMed=12477932;
 RA Srausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 [3]
 RN FUNCTION, AND INTERACTION WITH SP1.
 RP MEDLINE=22129221; PubMed=12134007;
 RA Takada N., Sanda T., Okamoto H., Yang J.-P., Asamitsu K., Sarol L.,
 RA Kimura G., Uranishi H., Tetsuka T., Okamoto T.;
 RT "RelA-associated inhibitor blocks transcription of human
 RT immunodeficiency virus type 1 by inhibiting NF-kappaB and Sp1
 RT actions";
 RL J. Virol. 76:8019-8030 (2002).
 [4]
 RN FUNCTION, DISEASE, AND INTERACTION WITH TP53.
 RP MEDLINE=22447931; PubMed=12524540;
 RA Bergamaschi D., Samuels Y., O'Neil N.J., Trigiante G., Crook T.,
 RA Hsieh J.-K., O'Connor D.J., Zhong S., Campargue I., Tomlinson M.L.,
 RA Kuwabara P.E., Lu X.;
 RT "RelA-associated protein is a key inhibitor of p53 conserved from worm to
 RT human";
 RL Nat. Genet. 33:162-167 (2003).
 CC -!- FUNCTION: Regulator that plays a central role in regulation of
 CC apoptosis and transcription via its interaction with NF-kappa-B

CC and p53/TP53 proteins. Blocks transcription of HIV-1 virus by
 CC inhibiting the action of both NF-kappa-B and Sp1. Also inhibits
 CC TP53 function, possibly by preventing the association between TP53
 CC and ASPP1 or ASPP2, and therefore suppressing the subsequent
 CC activation of apoptosis.
 CC -!- SUBUNIT: Interacts with RELA NF-kappa-B subunit and with Sp1 via
 CC its C-terminus part. Interacts with TP53.
 CC -!- SUBCELLULAR LOCATION: Predominantly nuclear.
 CC -!- TISSUE SPECIFICITY: Highly expressed in heart, placenta and
 CC prostate. Weakly expressed in brain, liver, skeletal muscle,
 CC testis and peripheral blood leukocyte.
 CC -!- DISEASE: Defects in RAI may be a cause of breast cancers. It is
 CC overexpressed in many patients suffering from breast carcinomas
 CC and expressing a wild-type TP53 protein.
 CC -!- SIMILARITY: Belongs to the ASPP family.
 CC -!- SIMILARITY: Contains 2 ANK repeats.
 CC -!- SIMILARITY: Contains 1 SH3 domain.
 CC -!- CAUTION: Ref.1 sequences differ from that shown due to
 CC frameshifts in positions 5, 61, 141 and 170.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL; AF078036; AAD27004.1; ALT_FRAME.
 DR EMBL; AF078037; AAD27005.1; ALT_FRAME.
 DR EMBL; BC020589; AAH20589.1; -.
 DR HSSP; Q13625; 1YCS.
 DR MIM; 607463; -.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF000023; ank; 2.
 DR Pfam; PF000018; SH3; 1.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR ProDom; PD000066; SH3; 1.
 DR SMART; SM00248; ANK; 2.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS00088; ANK_REPEAT; 2.
 DR PROSITE; PS00297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00002; SH3; 1.
 DR Apoptosis; Transcription regulation; Repressor; Nuclear protein;
 KW Repeat; ANK repeat; SH3 domain.
 FT REPEAT 238 270 ANK 1.
 FT REPEAT 271 303 ANK 2.
 FT DOMAIN 337 399 SH3.
 FT DOMAIN 32 181 PRO-RICH.
 FT SSSEQUENCE 407 AA; 44105 MW; 2D281AD004F352EF CRC64;
 SQ

Query Match 50.7%; Score 37; DB 1; Length 407;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 3;
 QY 6 WMA 9
 DB 377 WMA 380

RESULT 66
 PTK6_HUMAN
 ID PTK6_HUMAN STANDARD; PRT; 451 AA.
 AC Q13852;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Tyrosine-protein kinase 6 (BC 2.7.1.12) (Breast tumor kinase)
 DE (Tyrosine-protein kinase brk).
 GN PTK6 OR BRK.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN NCBI_TaxID=9606;
 RP [1]
 RC SEQUENCE FROM N.A.
 RX TISSUE=Brain; tumor;
 RA MEDLINE=94309916; PubMed=8036022;
 RA Mitchell P.J., Barker K.T., Martindale J.E., Kamalati T., Lowe P.N.,
 RA Page M.J., Gusterson B.A., Crompton M.R.;
 RT "Cloning and characterisation of cDNAs encoding a novel non-receptor
 RT tyrosine kinase, brk, expressed in human breast tumours";
 RL Oncogene 9:2383-2390(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Melanocyte;
 RA MEDLINE=98411955; PubMed=9749526;
 RA Lee H., Kim M., Lee K.-H., Kang K.-N., Lee S.-T.;
 RT "Exon-intron structure of the human PTK6 gene demonstrates that PTK6
 RT constitutes a distinct family of non-receptor tyrosine kinase";
 RL Mol. Cells 8:401-407(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Sabbage A.K., Bagdikian C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.V., Cleve C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Grahame D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Hackle B., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Letavajala M.H., Leverhulme M.A., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.B., McConachie L.J., McEay K., McKersin T.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkhan R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromas A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20";
 RL Nature 414:865-871(2001).
 RN [4]
 RP CHARACTERIZATION, AND MUTAGENESIS OF LYS-219 AND TYR-447.
 RX MEDLINE=97394849; PubMed=8940083;
 RA Kamalati T., Jolin H.E., Michell P.J., Barker K.T., Jackson L.E.,
 RA Dean C.J., Page M.J., Gusterson B.A., Crompton M.R.;
 RT "Brk, a breast tumor-derived non-receptor protein-tyrosine kinase,
 RT sensitizes mammary epithelial cells to epidermal growth factor";
 RL J. Biol. Chem. 271:30956-30963(1996).
 CC -!- FUNCTION: May function as an intracellular signal transducer in
 CC epithelial tissues. Overexpression in mammary cells leads to
 CC mitogenically sensitization to EGF, and results in a partially
 CC transformed phenotype.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -!- TISSUE SPECIFICITY: Very high level in colon and high levels in
 CC small intestine and prostate, and low levels in some fetal
 CC tissues. Expressed at low level in some breast tumors, but not in
 CC normal breast. Also found in melanocytes. Not expressed in heart,
 CC brain, placenta, lung, liver, skeletal muscle, kidney and
 CC pancreas.
 CC -!- PTM: Autophosphorylated.
 CC -!- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES.
 CC BRK/PTK6/SIK SUBFAMILY.

CC -!- SIMILARITY: Contains 1 SH2 domain.
 CC -!- SIMILARITY: Contains 1 SH3 domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL; X78549; CAAS5295.1; -;
 CC EMBL; U61412; AAC34935.1; -;
 CC EMBL; U61406; AAC34935.1; JOINED.
 CC EMBL; U61407; AAC34935.1; JOINED.
 CC EMBL; U61408; AAC34935.1; JOINED.
 CC EMBL; U61409; AAC34935.1; JOINED.
 CC EMBL; U61410; AAC34935.1; JOINED.
 CC EMBL; U61411; AAC34935.1; JOINED.
 CC EMBL; AL121829; CAC15825.1; -;
 CC F1R; S49016; S49016.
 CC HSP; F11362; IFGK.
 CC Genew; HGNC:9617; PTK6.
 CC MIM; 602004; -;
 CC GO; GO:0004715; F:non-membrane spanning protein tyrosine kina. . . ; TAS.
 CC GO; GO:0004668; P:protein amino acid phosphorylation; TAS.
 CC InterPro; IPR000719; Prot_kinase.
 CC InterPro; IPR000980; SH2.
 CC InterPro; IPR001452; SH3.
 CC InterPro; IPR001245; Tyr_kinase.
 CC InterPro; IPR008266; Tyr_kinase_AS.
 CC Pfam; PF00059; Pkinase; 1.
 CC Pfam; PF00017; SH2; 1.
 CC Pfam; PF03018; SH3; 1.
 CC PRINTS; PR00401; SH2DOMAIN.
 CC PRINTS; PR00452; SH3DOMAIN.
 CC PRINTS; PR00109; TYRKINASE.
 CC ProDom; PD000301; Prot_kinase; 1.
 CC ProDom; PD000033; SH2; 1.
 CC ProDom; PD000066; SH3; 1.
 CC SMART; SM00252; SH2; 1.
 CC SMART; SM00326; SH3; 1.
 CC SMART; SM00219; TyKc; 1.
 CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 CC PROSITE; PS00001; SH2; 1.
 CC PROSITE; PS00002; SH3; 1.
 CC KX Transferase; Tyrosine-protein kinase; ATP-binding; SH2 domain;
 CC SH3 domain; Phosphorylation.
 CC FT DOMAIN 11 72 SH3.
 CC FT DOMAIN 76 170 SH2.
 CC FT DOMAIN 191 445 PROTEIN KINASE.
 CC FT NP_BIND 197 205 ATP (BY SIMILARITY).
 CC FT BINDING 219 219 ATP (BY SIMILARITY).
 CC FT ACT_SITE 312 312 BY SIMILARITY.
 CC FT MOD_RES 342 342 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 CC FT MOD_RES 447 447 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MUTAGEN 219 219 K->M: ABOLISHES KINASE ACTIVITY AND CELL
 CC TRANSFORMATION.
 CC FT Y->F: DECREASE IN TRANSFORMING POTENTIAL.
 CC SEQUENCE 451 AA; 51834 MW; CDCAC0EE24E1BD7 CRC64;
 CC
 CC Query Match 50.7%; Score 37; DB 1; Length 451;
 CC Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 CC Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 6 WWA 9
 CC DB 44 WWA 47
 CC
 CC RESULT 67

CBL1 ARATH
 ID CBL1 ARATH STANDARD; PRT; 452 AA.
 AC Q9SR77;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE COBRA-like protein 1 precursor.
 GN CBL1 OR AT3G02210 OR F14P3.14.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21316720; PubMed=11130713;
 RA Salanoubat M., Lemcke K., Rieger M., Ansong W., Unselid M.,
 RA Fartmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,
 RA Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
 RA De Simone V., Choisme N., Artiguenave F., Robert C., Brottier P.,
 RA Wincker P., Cattolico L., Weissbach J., Saurin W., Quetier F.,
 RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
 RA Wurmbach E., Drzonek H., Erle H., Jordan N., Bangert S.,
 RA Wiedemann R., Kranz H., Voes H., Holland R., Brandt P., Nyakatura G.,
 RA Vezzi A., D'Angelo M., Pallavicini A., Toppi S., Simonati B.,
 RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordsiek G.,
 RA Reichelt J., Scharfe M., Schoen O., Barges M., Terol J., Climent J.,
 RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
 RA Cooke R., Laudie M., Berger-Liauro C., Purnelle B., Masuy D.,
 RA De Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
 RA Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,
 RA Mauhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
 RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
 RA Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
 RA Cressy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
 RA Pal G., Millscher J., Sellers P., Gill J.E., Feldblyum T.V.,
 RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White C., Venter J.C.,
 RA Sasamoto S., Kimura T., Nakamura Y., Sato S., Kato T., Asamizu E.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Kawashima K., Kishida Y.,
 RA Nakayama S., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
 RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
 thaliana.";
 RL Nature 408:820-822 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
 RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
 RA Hayashizaki Y., Shinzaki K.;
 RT "Arabidopsis thaliana full-length cDNA.";
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=22954850; PubMed=14593172;
 RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
 RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
 RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
 RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
 RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Anesari Y.,
 RA Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
 RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
 RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
 RA Khan S., Koeseema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
 RA Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
 RA Satou M., Tamse R., Vaysberg E.K., Wong C., Yamamura Y.,
 RA Yuan S., Shinzaki K., Davis R.W., Theologis A., Ecker J.R.;
 RT "Empirical analysis of transcriptional activity in the Arabidopsis
 genome.";
 RL Science 302:842-846 (2003).

RN RP
 TX TISSUE SPECIFICITY.
 RA MEDLINE=22263984; PubMed=12376623;
 RA Roudier P., Schindelman G., Desalle R., Benfey P.M.;
 RT "The COBRA family of putative GPI-anchored proteins in Arabidopsis. A
 RT new fellowship in expansion.";
 RL Plant Physiol. 130:538-548 (2002).
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (potential).
 CC -!- TISSUE SPECIFICITY: Expressed in roots, stems, leaves, flowers and
 CC siliques.
 CC -!- SIMILARITY: Belongs to the COBRA family.
 CC [1]
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AC009755; AAF02128.1; --
 DR EMBL; AK118555; BAC43156.1; --
 DR EMBL; BT005373; AAO63437.1; --
 DR InterPro; IPR006918; Phytochel synth.
 DR Pfam; PF04833; phytochel synth; 1.
 KW Membrane, GPI-anchor, Glycoprotein, Signal, Lipoprotein.
 FT SIGNAL 1 33 POTENTIAL.
 FT CHAIN 34 432 COBRA-LIKE PROTEIN 1.
 FT PROPEP 433 452 REMOVED IN NATURE FORM (POTENTIAL).
 FT LIPID 432 432 GPI-anchor amidated serine (POTENTIAL).
 FT CARBOHYD 42 42 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 167 167 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 175 175 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 214 214 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 239 239 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 254 254 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 323 323 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 338 338 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 357 357 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 452 AA; 50363 MW; A1A820B0AF96227 CRC64;
 Query Match 50.7%; Score 37; DB 1; Length 452;
 Best Local Similarity 66.7%; Pred. No. 2.6e+02;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 WTNIWVWAK 10
 ||| ||||
 Db 77 WTLCNSWAK 85
 RESULT 68
 COBR ARATH STANDARD; PRT; 456 AA.
 ID COBR ARATH STANDARD; PRT; 456 AA.
 AC Q94KT8; O81811; O8LDZ4;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE COBRA protein precursor (Cell expansion protein).
 GN COB OR AT5G0920 OR MSL3.40.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.; TISSUE SPECIFICITY, AND SUBCELLULAR LOCATION.
 RX MEDLINE=21231189; PubMed=11331607;
 RA Schindelman G., Morikami A., Jung J., Baskin T.I., Carpita N.C.,
 RA Derbyshire P., McCann M.C., Benfey P.N.;
 RT "COBRA encodes a putative GPI-anchored protein, which is polarly
 RT localized and necessary for oriented cell expansion in Arabidopsis.";
 RL Genes Dev. 15:1115-1127 (2001).

[2]
 CC SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RX MEDLINE=98162728; PubMed=9501997;
 RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.
 RT Sequence features of the regions of 1,191,918 bp covered by seventeen
 RT physically assigned p1 clones.";
 RL DNA Res. 4:401-414(1997).
 [3]
 CC SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RX MEDLINE=22954850; PubMed=14593172;
 RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
 RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
 RA Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
 RA Miranda M., Quach H.B., Tripp M., Chong C.H., Lee J.M., Tortum M.J.,
 RA Arakawa T., Barni C., Banno F., Bower J., Brooks S.Y., Carninci P.,
 RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
 RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
 RA Khan S., Koesema E., Iehida J., Jiang P.X., Jones T., Kawai J.,
 RA Kaniya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
 RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,
 RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
 RT "Empirical analysis of transcriptional activity in the Arabidopsis
 RT genome.";
 RL Science 302:842-846(2003).
 [4]
 CC SEQUENCE FROM N.A.
 RP SEQUENCE FROM N.A.
 RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
 RA Feldman K.A.;
 RT "Full-length cDNA from Arabidopsis thaliana."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 [5]
 CC SEQUENCE OF 62-456 FROM N.A.
 RA Leuchter R., Wolf K., Zimmermann M.;
 RT "Isolation of an Arabidopsis thaliana cDNA complementing a
 RT Schizosaccharomyces pombe mutant which is deficient in phytochelatin
 RT synthesis."
 RL (In) Plant Gene Register PGR98-147.
 [6]
 CC TISSUE SPECIFICITY.
 RX MEDLINE=22263984; PubMed=12376623;
 RA Roudier F., Schindelman G., DeSalle R., Benfey P.N.;
 RT "The COBRA family of putative GPI-anchored proteins in Arabidopsis. A
 RT new fellowship in expansion.";
 RL Plant Physiol. 130:538-548(2002).
 CC -!- FUNCTION: Involved in determining the orientation of cell
 CC expansion, probably by playing an important role in cellulose
 CC deposition. May act by recruiting cellulose synthesizing complexes
 CC to discrete positions on the cell surface.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
 CC (probable). Located on the longitudinal sides of the cell rather
 CC than on the apical or basal sides.
 CC -!- TISSUE SPECIFICITY: Expressed in roots, stems, leaves, flowers and
 CC siliques. Upregulated in the root zone of rapid longitudinal
 CC expansion.
 CC -!- MISCELLANEOUS: A partial protein missing the N-terminal signal
 CC peptide was reported to complement a yeast mutant defective in
 CC phytochelatin synthesis. It is therefore possible that COBRA binds
 CC divalent metals.
 CC -!- SIMILARITY: Belongs to the COBRA family.
 CC -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
 CC gene model prediction.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch.

 CC EMBL; AF319663; AAK56072.1; -
 CC EMBL; AB008269; BAB10641.1; ALT_SEQ.
 CC EMBL; AY094402; AAM19781.1; -
 CC EMBL; BT001119; AAN64510.1; -
 CC DR EMBL; AY085712; AAM62930.1; -
 CC DR EMBL; AJ005787; CAA07251.1; ALT_INIT.
 CC FIR; TS2038; TS2038.
 CC DR InterPro; IPR006918; Phytochel synth.
 CC Pfam; PF04833; phytochel synth; 1.
 KW Membrane; GPI-anchor; Glycoprotein; Signal; Lipoprotein.
 FT SIGNAL 1 36
 FT CHAIN 37 431
 FT PROPEP 432 456
 FT LIPID 431 431
 FT CARBOHYD 45 45
 FT CARBOHYD 170 170
 FT CARBOHYD 178 178
 FT CARBOHYD 217 217
 FT CARBOHYD 242 242
 FT CARBOHYD 258 258
 FT CARBOHYD 328 328
 FT CARBOHYD 343 343
 FT CARBOHYD 362 362
 FT CARBOHYD 456 456
 FT SEQUENCE 456 AA; 51202 MW; 986840C2E76D2C3 CRC64;
 Query Match 50.7%; Score 37; DB 1; Length 456;
 Best Local Similarity 66.7%; Pred. No. 2.6e+02;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 WTNIWWAK 10
 DB 80 WTLGKWKAK 88

 CC RESULT 69
 CC DPG2 MOUSE
 ID DPG2 MOUSE STANDARD; PRT; 459 AA.
 AC Q9QZM2; O35614;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE DNA polymerase gamma subunit 2, mitochondrial precursor (SC 2.7.7.7)
 DE (Mitochondrial DNA polymerase accessory subunit) (PoIG-beta) (McPoIB)
 DE (DNA polymerase gamma accessory 55 kDa subunit) (p55).
 GN POLG2 OR MTPOLB.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Swiss Webster / NIH;
 RX MEDLINE=20133043; PubMed=10666468;
 RA Carrodegas J.A., Bogenhagen D.F.;
 RT "Protein sequences conserved in prokaryotic aminoacyl-tRNA synthetases
 RT are important for the activity of the processivity factor of human
 RT mitochondrial DNA polymerase."
 RL Nucleic Acids Res. 28:1237-1244(2000).
 [2]
 CC SEQUENCE OF 164-459 FROM N.A.
 RA Kaguni L.S.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 CC X-RAY CRYSTALLOGRAPHY (1.95 ANGSTROMS) OF 1-459.
 RX MEDLINE=21111065; PubMed=1172710;
 RA Carrodegas J.A., Theis K., Bogenhagen D.F., Kisker C.;
 RT "Crystal structure and deletion analysis show that the accessory
 RT subunit of mammalian DNA polymerase gamma, Pol gamma B, functions as
 RT a homodimer."
 RL Mol. Cell 7:43-54(2001).

CC -/- FUNCTION: Mitochondrial polymerase processivity subunit.
 CC Stimulates the polymerase and exonuclease activities, and
 CC increases the processivity of the enzyme. Binds to ss-DNA.
 CC + [DNA] (N).
 CC -/- SUBUNIT: Heterotrimer composed of a catalytic subunit and an
 CC homodimer of accessory subunits.
 CC -/- SUBCELLULAR LOCATION: Mitochondrial.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AF177202; RAD56641.1; -;
 DR EMBL: AF006972; AAB62894.1; -;
 DR PDB: 1GSH; 14-MAR-01.
 DR PDB: 1GSI; 14-MAR-01.
 DR MGD: MGI:1354947; Polq2.
 DR GO: GO:0003895; F:gamma DNA-directed DNA polymerase activity; IDA.
 DR InterPro: IPR004154; HGTP_anticoodon; 1.
 DR Pfam: PF03129; HGTP_anticoodon; 1.
 KW Transferase; DNA-directed DNA polymerase; DNA replication;
 KW DNA-binding; Mitochondrion; Transit peptide; 3D-structure.
 FT TRANSIT 1 ?
 FT CHAIN 1 ? MITOCHONDRION (POTENTIAL).
 FT TRAIIN 1 459 DNA POLYMERASE GAMMA SUBUNIT 2.
 FT CONFLICT 226 226 L -> S (IN REF. 2).
 SQ SEQUENCE 459 AA; 51432 MW; D4SB5164245DBICQ CRC64;

Query Match 50.7%; Score 37; DB 1; Length 459;
 Best Local Similarity 31.2%; Pred. No. 2.6e+02;
 Matches 5; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

Qy 1 EWTNIV-----WVAK 10
 Db 224 QWDFWLRHLLWVFX 239

RESULT 70
 NU4M SCYCA
 ID NU4M SCYCA STANDARD; PRT; 460 AA.
 AC 079410;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE NADH-ubiquinone oxidoreductase chain 4 (EC 1.6.5.3).
 GN MTND4 OR ND4 OR NAD4 OR NAD4 OR NAD4.
 OS Scyllorhinus canalicula (Spotted dogfish) (Spotted catshark).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
 OC Elasmobranchii; Galeomorphii; Galeoidea; Carchariniiformes;
 OC Scyllorhinidae; Scyllorhinus.
 OX NCBI_TaxID=7830;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RX MEDLINE=98393590; PubMed=9725850;
 RA Delarbre C., Spruyt N., Delmarre C., Gallut C., Barriel V.,
 RA Janvier P., Laudet V., Gachelin G.;
 RT "The complete nucleotide sequence of the mitochondrial DNA of the
 RT dogfish, Scyllorhinus canalicula.";
 RL Genetics 150:331-344 (1998).
 CC -/- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC or send an email to license@isb-sib.ch.
 CC
 DR EMBL: Y16067; CAA76028.1; -;
 DR PIR: T11309; T11309.
 DR InterPro: IPR003918; NADHub_oxred4.
 DR InterPro: IPR001750; Oxidored_g1.
 DR InterPro: IPR000260; Oxidored_g5_N.
 DR Pfam: PF00361; oxidored_g1; 1.
 DR Pfam: PF01059; oxidored_g5_N; 1.
 DR PRINTS: PR01437; NUOXDRDTASE4.
 KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
 SQ SEQUENCE 460 AA; 52148 MW; 44E7A8A528AFE366 CRC64;

Query Match 50.7%; Score 37; DB 1; Length 460;
 Best Local Similarity 57.1%; Pred. No. 2.6e+02;
 Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 WTNIVWW 8
 Db 191 WAKFWW 197

RESULT 71
 KCS2 HUMAN
 ID KCS2 HUMAN STANDARD; PRT; 477 AA.
 AC Q9ULS6;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Potassium voltage-gated channel subfamily S member 2 (Potassium
 DE channel Kv9.2) (Delayed-rectifier K+ channel alpha subunit 2).
 GN KCS2 OR KIAA1144.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCES FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20039618; PubMed=10574461;
 RA Hirosewa M., Nagase T., Ishikawa K.-I., Kikuno R., Nomura N.,
 RA Ohara O.;
 RT "Characterization of cDNA clones selected by the Genemark analysis
 RT from size-fractionated cDNA libraries from human brain.";
 RL DNA Res. 6:329-336 (1999).
 [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Peingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udgin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullen S.J.,
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.W., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 CC -/- FUNCTION: Potassium channel subunit. Modulates channel activity
 CC and reduces the ion flow (By similarity).
 CC -/- SUBUNIT: Heteromultimer with KCNB1 and with KCNB2. Does not form
 CC homomultimers. Might also bind to other channel proteins (By

CC similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. May not reach the
 CC plasma membrane but remain in an intracellular compartment in the
 CC absence of KCNB1 (By similarity).
 CC -!- DOMAIN: The segment S4 is probably the voltage-sensor and is
 CC characterized by a series of positively charged amino acids at
 CC every third position.
 CC -!- SIMILARITY: Belongs to the potassium channel family. S subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AB032970; BAA6458.1; ALC_INIT.
 CC EMBL; BC027932; AAH27932.1; -.
 CC EMBL; BC034778; AAH34778.1; -.
 CC Genew; HGNC:6301; KCNS2.
 CC MIM; 602906; -.
 CC InterPro; IPR000210; BVB_POZ.
 CC InterPro; IPR003821; Ion_trans.
 CC InterPro; IPR001622; K+channel_pore.
 CC InterPro; IPR003091; K_channel.
 CC InterPro; IPR003131; K_tetra.
 CC InterPro; IPR003971; Kv9_channel.
 CC InterPro; IPR003968; Kv_channel.
 CC InterPro; IPR005820; M+channel_nlg.
 CC Pfam; PF00520; ion_trans; 1.
 CC Pfam; PF02214; K_tetra; 1.
 CC PRINTS; PR00169; KCHANNEL.
 CC PRINTS; PR01494; KV9CHANNEL.
 CC PRINTS; PR01491; KVCHANNEL.
 CC SMART; SMC0225; BTB; 1.
 CC Transport; Ion transport; Ionic channel, Voltage-gated channel;
 CC Potassium channel; Potassium; Potassium transport; Transmembrane;
 CC Multigene family.
 CC DOMAIN 1 187
 CC TRANSMEM 188 208 CYTOPLASMIC (POTENTIAL).
 CC TRANSMEM 233 253 SEGMENT S1 (POTENTIAL).
 CC DOMAIN 254 261 SEGMENT S2 (POTENTIAL).
 CC TRANSMEM 262 282 CYTOPLASMIC (POTENTIAL).
 CC TRANSMEM 293 313 SEGMENT S3 (POTENTIAL).
 CC DOMAIN 314 329 SEGMENT S4 (POTENTIAL).
 CC TRANSMEM 330 350 CYTOPLASMIC (POTENTIAL).
 CC TRANSMEM 362 382 SEGMENT S5 (POTENTIAL).
 CC DOMAIN 390 410 SEGMENT S6 (POTENTIAL).
 CC TRANSMEM 411 477 CYTOPLASMIC (POTENTIAL).
 CC SEQUENCE 477 AA; 54237 MW; P231AD99EC02EB46 CRC64;
 CC
 CC Query Xatch 50.7%; Score 37; DB 1; Length 477;
 CC Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 CC Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC QY 6 WWA 9
 CC DB 365 WWA 368
 CC
 CC RESULT 72
 CC KCS2_MOUSE
 CC ID KCS2_MOUSE STANDARD; PRT; 477 AA.
 CC AC Q35174;
 CC DT 28-FEB-2003 (Rel. 41, Created)
 CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
 CC DT 15-MAR-2004 (Rel. 43, Last annotation update)
 CC DE Potassium voltage-gated channel subfamily S member 2 (Potassium
 CC channel Kv9.2) [Delayed-rectifier K+ channel alpha subunit 2].
 CC GN KCNS2.
 CC OS Mus musculus (Mouse).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10390;
 RN [1]
 RP SEQUENCE FROM N.A., AND FUNCTION.
 RX MEDLINE=97450962; PubMed=9305895;
 RA Salinas M., Duprat F., Heurteaux C., Hugnot J.-P., Lazdunski M.;
 RT "New modulatory alpha subunits for mammalian shab K+ channels";
 RL J. Biol. Chem. 272:24371-24379 (1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CS7BL/6; TISSUE=Brain;
 RX MEDLINE=2388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mollany S.J.,
 RA Rosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Villalón D.K., Muzny D.M., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 CC -!- FUNCTION: Potassium channel subunit. Modulates channel activity
 CC and reduces the ion flow.
 CC -!- SUBUNIT: Heteromultimer with KCNB1 and with KCNB2. Does not form
 CC homooligomers. Might also bind to other channel proteins.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. May not reach the
 CC plasma membrane but remain in an intracellular compartment in the
 CC absence of KCNB1.
 CC -!- TISSUE SPECIFICITY: Detected in brain, but not in the other
 CC tissues tested. Expression was highest in the olfactory bulb,
 CC cerebral cortex, hippocampus, habenula, basolateral amygdaloid
 CC nuclei and cerebellum.
 CC -!- DOMAIN: The segment S4 is probably the voltage-sensor and is
 CC characterized by a series of positively charged amino acids at
 CC every third position.
 CC -!- SIMILARITY: Belongs to the potassium channel family. S subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF008574; AAB72051.1; -.
 CC EMBL; BC059833; AAB59833.1; -.
 CC MGD; MGI:1197011; Kcns2.
 CC InterPro; IPR000210; BTB_POZ.
 CC InterPro; IPR005821; Ion_trans.
 CC InterPro; IPR001622; K+channel_pore.
 CC InterPro; IPR003091; K_channel.
 CC InterPro; IPR003131; K_tetra.
 CC InterPro; IPR003971; Kv9_channel.
 CC InterPro; IPR003968; Kv_channel.
 CC Pfam; PF00520; ion_trans; 1.
 CC Pfam; PF02214; K_tetra; 1.
 CC PRINTS; PR00169; KCHANNEL.
 CC PRINTS; PR01494; KV9CHANNEL.
 CC PRINTS; PR01491; KVCHANNEL.
 CC SMART; SMC0225; BTB; 1.

Transport; Ion transport; Ionic channel; Voltage-gated channel;
 KW Potassium channel; Potassium; Potassium transport; Transmembrane;
 KW Multigene family.

FT DOMAIN 1 187 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 188 208 SEGMENT S1 (POTENTIAL).
 FT TRANSMEM 233 253 SEGMENT S2 (POTENTIAL).
 FT DOMAIN 254 261 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 262 282 SEGMENT S3 (POTENTIAL).
 FT TRANSMEM 293 313 SEGMENT S4 (POTENTIAL).
 FT DOMAIN 314 328 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 329 349 SEGMENT S5 (POTENTIAL).
 FT DOMAIN 362 382 SEGMENT H5 (PORE-FORMING) (POTENTIAL).
 FT TRANSMEM 390 410 SEGMENT S6 (POTENTIAL).
 FT DOMAIN 411 477 CYTOPLASMIC (POTENTIAL).
 SQ SEQUENCE 477 AA; 54289 YW; C7AD7AA3AE312B2C CRC64;

Query Match 50.7%; Score 37; DB 1; Length 477;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 WWA 9
 ||||
 Db 365 WWA 368

RESULT 73
 KCS2 RAT
 ID KCS2 RAT STANDARD; PRT; 477 AA.
 AC Q9ER26;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Potassium voltage-gated channel subfamily S member 2 (Potassium
 channel Kv9.2) (Delayed-rectifier K⁺ channel alpha subunit 2).
 DE channel Kv9.2) (Delayed-rectifier K⁺ channel alpha subunit 2).
 GN KCNS2.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wistar; TISSUE=Brain;
 RA MEDLINE=21889041; PubMed=11891605;
 RX Davies A.R., Kozlowski R.Z.;
 RT "Kv channel subunit expression in rat pulmonary arteries.";
 RL Lung 179:147-161(2001).
 CC -!- FUNCTION: Potassium channel subunit. Modulates channel activity
 CC and reduces the ion flow (By similarity).
 CC -!- SUBUNIT: Heteromultimer with KCNB1 and with KCNB2. Does not form
 CC homomultimers. Might also bind to other channel proteins (By
 CC similarity).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. May not reach the
 CC plasma membrane but remain in an intracellular compartment in the
 CC absence of KCNB1 (By similarity).
 CC -!- TISSUE SPECIFICITY: Detected in brain, lung and in pulmonary
 CC arteries.
 CC -!- DOMAIN: The segment S4 is probably the voltage-sensor and is
 CC characterized by a series of positively charged amino acids at
 CC every third position.
 CC -!- SIMILARITY: Belongs to the potassium channel family. S subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AJ296090; CAC14912.1; -
 DR InterPro; IPR000210; EBF_POZ.
 DR InterPro; IPR005821; Ion_trans.
 DR InterPro; IPR001622; K+channel_pore.

InterPro; IPR003091; K channel.
 DR InterPro; IPR003131; K tetra.
 DR InterPro; IPR003971; Kv9 channel.
 DR InterPro; IPR003968; Kv9 channel.
 DR InterPro; IPR005820; M-channel_nlg.
 DR Pfam; PF00520; ion_trans; 1.
 DR Pfam; PF02214; K_tetra; 1.
 DR PRINTS; PR00169; KCHANNEL.
 DR PRINTS; PR01494; KV9CHANNEL.
 DR PRINTS; PR01491; KVCHANNEL.
 DR SMART; SM00225; BTB; 1.
 KW Transport; Ion transport; Ionic channel; Voltage-gated channel;
 KW Potassium channel; Potassium; Potassium transport; Transmembrane;
 KW Multigene family.

FT DOMAIN 1 187 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 188 208 SEGMENT S1 (POTENTIAL).
 FT TRANSMEM 233 253 SEGMENT S2 (POTENTIAL).
 FT DOMAIN 254 262 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 263 282 SEGMENT S3 (POTENTIAL).
 FT TRANSMEM 293 313 SEGMENT S4 (POTENTIAL).
 FT DOMAIN 314 328 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 329 349 SEGMENT S5 (POTENTIAL).
 FT DOMAIN 362 382 SEGMENT H5 (PORE-FORMING) (POTENTIAL).
 FT TRANSMEM 390 410 SEGMENT S6 (POTENTIAL).
 FT DOMAIN 411 477 CYTOPLASMIC (POTENTIAL).
 SQ SEQUENCE 477 AA; 54317 MW; 23199B1BB9C5D5C CRC64;

Query Match 50.7%; Score 37; DB 1; Length 477;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 WWA 9
 ||||
 Db 365 WWA 368

RESULT 74
 DPG2 HUMAN
 ID DPG2 HUMAN STANDARD; PRT; 485 AA.
 AC Q9UHN1; O00419; Q9UK35; Q9UK94;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE DNA polymerase gamma subunit 2, mitochondrial precursor (EC 2.7.7.7)
 DE (Mitochondrial DNA polymerase accessory subunit) (PolG-beta) (McpolB)
 DE (DNA polymerase gamma accessory 55 kDa subunit) (p55).
 GN POLG2 OR MTPOLB.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 CC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20133043; PubMed=10666468;
 RA Carrodegas J.A., Bogenhagen D.F.;
 RT "Protein sequences conserved in prokaryotic aminoacyl-tRNA synthetases
 RT are important for the activity of the processivity factor of human
 RT mitochondrial DNA polymerase.";
 RL Nucleic Acids Res. 28:1237-1244(2000).
 CC [2]
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cerebellum;
 RX MEDLINE=20076499; PubMed=10608893;
 RA Lim S.E., Longley M.J., Copeland W.C.;
 RT "The mitochondrial p55 accessory subunit of human DNA polymerase
 RT gamma enhances DNA binding, promotes processive DNA synthesis, and
 RT confers N-ethylmaleimide resistance.";
 RL J. Biol. Chem. 274:38197-38203(1999).
 CC [3]
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Johnson A.A., Teai Y.-C., Graves S.W., Johnson K.A.;
 RT "Human mitochondrial DNA polymerase holoenzyme: reconstitution and
 RT characterization."

Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

[4]

Sequence of 114-485 from N.A.

Medline=97298065; PubMed=9153213;

Wang Y., Farr C.L., Kaguni L.S.;

*Accessory subunit of mitochondrial DNA polymerase from Drosophila embryos. Cloning, molecular analysis, and association in the native enzyme.;

J. Biol. Chem. 272:13640-13646 (1997).

!- FUNCTION: Mitochondrial polymerase processivity subunit. Increases the polymerase and exonuclease activities, and stimulates the processivity of the enzyme. Binds to ss-DNA.

!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate + {DNA}(N).

!- SUBUNIT: Heterotrimer composed of a catalytic subunit and an homodimer of accessory subunits.

!- SUBCELLULAR LOCATION: Mitochondrial.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF142992; AAD50382.1; -

EMBL; AF177201; AAD56640.1; -

EMBL; AF184344; AAD56542.1; -

EMBL; U94703; AAC53321.1; -

HSRP; Q9QZM2; IGSH.

Genew; HGNC:9180; POLG2.

MIM; 604983; -

GO; GO:0005739; C:mitochondrion; NAS.

GO; GO:0003887; F:DNA-directed DNA polymerase activity; NAS.

GO; GO:0003497; F:single-stranded DNA binding; NAS.

GO; GO:0005281; P:DNA repair; NAS.

GO; GO:0006260; P:DNA replication; NAS.

InterPro; IPR004154; HGTP anticodon; 1.

Pfam; PF01129; HGTP anticodon; 1.

Transferase; DNA-directed DNA polymerase; DNA replication; DNA-binding; Mitochondrion; Transit peptide.

TRANSIT ?

CHAIN ?

MITOCHONDRION (POTENTIAL).

DNA POLYMERASE GAMMA SUBUNIT 2.

WTSVVFREQ -> MVDLGGVHGA (IN REF. 4).

R -> T (IN REF. 3).

S -> S (IN REF. 3 AND 4).

A -> T (IN REF. 1).

NKLYN -> TNFTTI (IN REF. 4).

SEQUENCE 485 AA; 54911 MW; B99734BFEA249192 CRC64;

Query Match 50.7%; Score 37; DB 1; Length 485;

Best Local Similarity 31.2%; Pred. No. 2.7e+02;

Matches 5; Conservative 2; Mismatches 3; Indels 6; Gaps 1;

QY 1 EWTNIV-----WWAK 10

DB 250 QWLDLFWLRHRLQWWRK 265

RESULT 75

KCS3_HUMAN

ID KCS3_HUMAN STANDARD; PRT; 491 AA.

AC Q9BQ31; O43651; O96B56;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Potassium voltage-gated channel subfamily S member 3 (Potassium channel Kv9.3) (Delayed-rectifier K+ channel alpha subunit 3).

GN KCS3.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A., AND VARIANT THR-450.

RC TISSUE=Lens epithelium;

RX MEDLINE=99413882; PubMed=10484328;

RA Shepard A.R., Rae J.L.;

RT "Electrically silent potassium channel subunits from human lens enzyme.;"

RT epithelium.;"

RL Am. J. Physiol. 277:C412-C424 (1999).

[2]

SEQUENCE FROM N.A.

RC TISSUE=Kidney, and Skin;

RX MEDLINE=2388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E., Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Frange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J., Bock S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.;"

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

!- FUNCTION: Potassium channel subunit. Modulates channel activity and reduces the ion flow (By similarity).

!- SUBUNIT: Heteromultimer with KCNB1 and with KCNB2. Does not form homomultimers. Might also bind to other channel proteins (By similarity).

!- SUBCELLULAR LOCATION: Integral membrane protein. May not reach the plasma membrane but remain in an intracellular compartment in the absence of KCNB1.

!- TISSUE SPECIFICITY: Detected in most tissues, but not in peripheral blood lymphocytes. The highest levels of expression are in lung.

!- DOMAIN: The segment S4 is probably the voltage-sensor and is characterized by a series of positively charged amino acids at every third position.

!- SIMILARITY: Belongs to the potassium channel family. S subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF043472; AAC13164.1; -

EMBL; BC004148; AAH04148.1; -

EMBL; BC004987; AAH04987.1; -

EMBL; BC015947; AAH15947.1; -

Genew; HGNC:6302; KCNS3.

MIM; 603888; -

InterPro; IPR000210; BTB_POZ.

InterPro; IPR005821; Ion trans.

InterPro; IPR001622; K+channel_pore.

InterPro; IPR003091; K_channel.

InterPro; IPR003131; K_tetra.

InterPro; IPR003971; Kv9_channel.

InterPro; IPR003968; Kv_channel.

InterPro; IPR005820; M+channel_nlg.

Pfam; PF00520; ion trans; 1.

Pfam; PF02214; K_tetra; 1.

DR PRINTS; PR0169; KCHANNEL.
 DR PRINTS; PR01494; KV9CHANNEL.
 DR PRINTS; PR01491; KVCHANNEL.
 DR SMART; SMC0225; EFB; 1.
 KW Transport; Ion transport; Ionic channel; Voltage-gated channel;
 KW Potassium channel; Potassium; Potassium transport; Transmembrane;
 KW Multigene family; Polymorphism.
 FT DOMAIN 1 185
 FT TRANSMEM 186 206
 FT TRANSMEM 219 239
 FT DOMAIN 240 253
 FT TRANSMEM 254 274
 FT TRANSMEM 288 308
 FT DOMAIN 309 323
 FT TRANSMEM 324 344
 FT DOMAIN 358 378
 FT TRANSMEM 386 406
 FT DOMAIN 407 491
 FT VARIANT 450 450
 /FTIG=VAR 014200.
 A -> T.
 S -> Y (IN REF. 2; AAH15947).
 L -> V (IN REF. 1 AND 2; AAH15947).
 SQ SEQUENCE 491 AA; 55984 MW; ASF12BF077A50DAD CRC64;

Query Match 50.7%; Score 37; DB 1; Length 491;
 Best Local Similarity 100.0%; Pred. No. 2.8e+02;
 Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 WWA 9
 ||||
 Db 361 WWA 364

Search completed: June 9, 2004, 18:09:12
 Job time : 38 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2004, 18:07:01 : Search time 39 Seconds

(without alignments)

80.902 Million cell updates/sec

Title: US-10-726-148a-15_COPY_428_437

Perfect score: 73

Sequence: 1 EWTNIWVWAK 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

SPTREMBL_25.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phage.*

10: sp_plant.*

11: sp_rudent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_virus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47.5	65.1	328	16 Q87G63	Q87G63 vibrio para
2	47	64.4	340	8 Q85FP6	Q85FP6 cyanidiosch
3	46	63.0	458	8 Q94RC3	Q94RC3 chimera mo
4	45	61.6	122	15 Q90DP7	Q90DP7 human immun
5	45	61.6	168	16 Q8ZQ42	Q8ZQ42 salmonella
6	45	61.6	168	16 Q8Z7P2	Q8Z7P2 salmonella
7	45	61.6	211	15 Q8AU06	Q8AU06 human immun
8	45	61.6	222	12 Q64901	Q64901 arctic grou
9	45	61.6	282	12 Q64900	Q64900 arctic grou
10	45	61.6	427	12 Q64899	Q64899 arctic grou
11	45	61.6	860	15 Q90QM3	Q90QM3 human immun
12	45	61.6	868	15 Q90VI2	Q90VI2 human immun
13	45	61.6	868	15 Q9YV00	Q9YV00 human immun
14	44	60.3	47	8 Q85FP6	Q85FP6 cyanidiosch
15	44	60.3	117	15 Q9YV26	Q9YV26 human immun
16	44	60.3	122	15 Q90DR7	Q90DR7 human immun

Q91JM9 human immun
Q8JEG8 human immun
Q95PA0 halocynthia
Q24539 drosophila
Q8M212 drosophila
Q8G7H3 human immun
Q8KU32 legioneella
Q69908 human immun
Q91GN0 arabidopsis
Q9GDK4 human immun
Q90G62 human immun
Q823W0 chlamydomphi
Q82MM4 streptomyce
Q8IVQ1 homo sapien
Q90VE9 human immun
Q9WUJ7 human immun
Q9WC60 human immun
Q9WC69 human immun
Q41564 human immun
Q8G870 human immun
Q9Y5C3 homo sapien
Q9Y5F7 homo sapien
Q9HAG6 homo sapien
Q9EX94 pseudomonas
Q9LDC3 ochromobactru
Q50329 escherichia
Q8S2U3 medicago tr
Q90DJ2 human immun
Q91JK9 human immun
Q8S2U2 medicago tr
Q8U19 human immun
Q8U18 human immun
Q41871 zea mays (m
Q9Z6J0 chlamydia p
Q8SCR7 pseudomonas
Q07171 mycobacteri
Q7U2T3 mycobacteri
Q8N8Z4 homo sapien
Q94C18 lycopersico
Q28654 archaeoglob
Q9EYR8 agrobacteri
Q8U738 agrobacteri
Q83W25 streptomyce
Q8ZF00 streptomyce
Q8HQM4 lepisosteus
Q8HMM5 lepisosteus
P94793 fusobacteri
Q9HQP5 halobacteri
Q90CW2 human immun
Q9DV12 human immun
Q8WQ32 leishmania
Q9DCV3 human immun
Q8XPQ3 ralstonia s
Q9W2J2 drosophila
Q86PC3 drosophila
Q86BGI drosophila
Q8A583 caulobacter
Q9Z4W1 streptomyce
Q7SVF0 human immun
Q33920 salmonella
Q7V4B3 prochloroco
Q9U3L9 caenorhabdi
Q9IE96 human immun
Q32398 rhodospseud
Q97XG1 sulfolobus
Q9UXL5 sulfolobus
Q96Z59 sulfolobus
Q9CD01 mycobacteri
Q06080 mycobacteri
Q8XR88 ralstonia s
Q9ZY36 raja radiat
Q8BCU9 human cytom
Q8AYW0 human cytom

90	41	56.2	501	16	Q8Y049	Q8Y049 ralstonia s
91	41	56.2	508	10	Q40411	Q40411 nicotiana p
92	41	56.2	523	16	Q8YX99	Q8YX99 anabaena sp
93	41	56.2	548	15	Q91ED6	Q91ED6 human immun
94	41	56.2	692	10	Q9SAH3	Q9SAH3 arabidopsis
95	41	56.2	764	10	Q9LMT9	Q9LMT9 arabidopsis
96	41	56.2	769	3	Q74253	Q74253 pycnoporus
97	41	56.2	835	2	Q9LC82	Q9LC82 aeromonas s
98	41	56.2	867	15	Q8Q7G8	Q8Q7G8 human immun
99	41	56.2	868	15	Q8UL87	Q8UL87 human immun
100	41	56.2	870	15	Q8UL81	Q8UL81 human immun
101	41	56.2	922	11	Q8K2D7	Q8K2D7 mus musculus
102	41	56.2	922	11	Q8BPR4	Q8BPR4 mus musculus
103	41	56.2	924	11	Q8BNE1	Q8BNE1 mus musculus
104	41	56.2	1213	16	Q7UNZ2	Q7UNZ2 rhodospirill
105	41	56.2	2321	16	Q8F5B9	Q8F5B9 leptospira
106	40.5	55.5	169	17	Q9Y9Q4	Q9Y9Q4 aeropyrum p
107	40.5	55.5	589	16	Q97JY8	Q97JY8 clostridium
108	40	54.8	91	10	Q7XK26	Q7XK26 oryza sativ
109	40	54.8	109	12	Q7TFD6	Q7TFD6 rhesus cyto
110	40	54.8	116	15	Q8UX00	Q8UX00 human immun
111	40	54.8	117	15	Q9YRR8	Q9YRR8 human immun
112	40	54.8	117	15	Q7SVH1	Q7SVH1 human immun
113	40	54.8	118	15	Q9ES55	Q9ES55 human immun
114	40	54.8	120	15	Q91HV8	Q91HV8 human immun
115	40	54.8	122	15	Q90DQ9	Q90DQ9 human immun
116	40	54.8	122	15	Q90DK8	Q90DK8 human immun
117	40	54.8	122	15	Q90DM1	Q90DM1 human immun
118	40	54.8	122	15	Q90DP5	Q90DP5 human immun
119	40	54.8	122	15	Q9YXN3	Q9YXN3 human immun
120	40	54.8	122	15	Q90DJ3	Q90DJ3 human immun
121	40	54.8	122	15	Q90DQ3	Q90DQ3 human immun
122	40	54.8	124	15	Q9Y203	Q9Y203 human immun
123	40	54.8	124	15	Q9Y203	Q9Y203 human immun
124	40	54.8	127	15	Q9Y203	Q9Y203 human immun
125	40	54.8	127	15	Q9Y206	Q9Y206 human immun
126	40	54.8	127	15	Q9Y211	Q9Y211 human immun
127	40	54.8	127	15	Q9Y211	Q9Y211 human immun
128	40	54.8	130	15	Q9Y211	Q9Y211 human immun
129	40	54.8	133	15	Q90641	Q90641 human immun
130	40	54.8	133	15	Q8UR77	Q8UR77 human immun
131	40	54.8	135	15	Q7SM23	Q7SM23 human immun
132	40	54.8	135	16	Q88C29	Q88C29 pseudomonas
133	40	54.8	135	16	Q87TZ9	Q87TZ9 pseudomonas
134	40	54.8	139	16	Q9KAF7	Q9KAF7 bacillus ha
135	40	54.8	139	16	Q9KAF7	Q9KAF7 bacillus ha
136	40	54.8	152	15	Q7SM41	Q7SM41 human immun
137	40	54.8	157	10	Q8GSF6	Q8GSF6 oryza sativ
138	40	54.8	164	11	Q8R5D4	Q8R5D4 mus musculus
139	40	54.8	164	15	Q913A7	Q913A7 human immun
140	40	54.8	166	16	Q8P7C7	Q8P7C7 xanthomonas
141	40	54.8	170	12	O11428	O11428 duck adenov
142	40	54.8	177	15	Q91EB0	Q91EB0 human immun
143	40	54.8	209	2	Q9S558	Q9S558 pseudomonas
144	40	54.8	218	10	Q9FGX9	Q9FGX9 arabidopsis
145	40	54.8	218	15	Q8JAM1	Q8JAM1 human immun
146	40	54.8	223	16	Q8ZSD5	Q8ZSD5 anabaena sp
147	40	54.8	236	3	Q9HEC7	Q9HEC7 neurospora
148	40	54.8	276	13	Q7TFD8	Q7TFD8 rhesus cyto
149	40	54.8	315	10	Q7XQK9	Q7XQK9 oryza sativ
150	40	54.8	318	16	Q9CJG6	Q9CJG6 pasteurilla
151	40	54.8	332	17	Q92922	Q92922 archaeoglob
152	40	54.8	333	17	Q9UZH2	Q9UZH2 pyrococcus
153	40	54.8	341	12	Q7F5J0	Q7F5J0 cryptophleb
154	40	54.8	350	16	Q9RZ91	Q9RZ91 deinococcus
155	40	54.8	353	3	Q0B980	Q0B980 saccharomyc
156	40	54.8	359	16	Q9X0M1	Q9X0M1 thermotoga
157	40	54.8	368	10	Q8RU25	Q8RU25 oryza sativ
158	40	54.8	371	16	O06741	O06741 bacillus su
159	40	54.8	375	16	Q9A935	Q9A935 caulobacter
160	40	54.8	400	10	Q8LJRI	Q8LJRI zea mays (m
161	40	54.8	405	2	Q9FIP9	Q9FIP9 burkholderi
162	40	54.8	405	2	Q9LAN5	Q9LAN5 burkholderi

163	40	54.8	426	16	Q8XSL4	Q8XSL4 ralstonia s
164	40	54.8	460	8	Q94TBC	Q94TBC neocapelus
165	40	54.8	491	17	O30220	O30220 archaeoglob
166	40	54.8	502	15	Q91EB3	Q91EB3 human immun
167	40	54.8	513	10	Q8S9C0	Q8S9C0 solanum tub
168	40	54.8	541	17	Q8U4G9	Q8U4G9 pyrococcus
169	40	54.8	544	17	O57945	O57945 pyrococcus
170	40	54.8	589	2	Q9Z4P9	Q9Z4P9 bacillus ci
171	40	54.8	591	16	Q9AA83	Q9AA83 caulobacter
172	40	54.8	739	12	Q99A03	Q99A03 tt virus. o
173	40	54.8	852	15	O56567	O56567 human immun
174	40	54.8	855	15	Q8ADM3	Q8ADM3 human immun
175	40	54.8	855	15	Q7SVL0	Q7SVL0 human immun
176	40	54.8	857	15	Q8AQS1	Q8AQS1 human immun
177	40	54.8	858	15	Q8JBT2	Q8JBT2 human immun
178	40	54.8	859	15	O11947	O11947 human immun
179	40	54.8	863	15	Q8Q7J0	Q8Q7J0 human immun
180	40	54.8	868	15	Q8JBV0	Q8JBV0 human immun
181	40	54.8	871	15	Q8Q7J2	Q8Q7J2 human immun
182	40	54.8	914	2	Q9R6Y9	Q9R6Y9 alteromonas
183	40	54.8	947	12	Q8B489	Q8B489 dengue viru
184	40	54.8	1063	10	Q8S9V3	Q8S9V3 oryza sativ
185	39.5	54.1	307	16	Q92X20	Q92X20 rhizobium m
186	39.5	54.1	397	16	Q8PTU1	Q8PTU1 xanthomonas
187	39	53.4	49	12	Q9DR80	Q9DR80 tt virus. o
188	39	53.4	88	16	Q8E817	Q8E817 streptomyce
189	39	53.4	94	9	Q8S9Q7	Q8S9Q7 bacterioph
190	39	53.4	105	6	Q95K82	Q95K82 macaca fasc
191	39	53.4	109	10	O8LNO1	O8LNO1 oryza sativ
192	39	53.4	111	10	Q7XILX7	Q7XILX7 oryza sativ
193	39	53.4	113	15	Q8UOX2	Q8UOX2 human immun
194	39	53.4	115	15	Q7ZJN8	Q7ZJN8 human immun
195	39	53.4	116	15	Q7ZJN9	Q7ZJN9 human immun
196	39	53.4	117	15	Q9Q6F2	Q9Q6F2 human immun
197	39	53.4	117	15	Q9Q6F3	Q9Q6F3 human immun
198	39	53.4	117	15	Q7SVF5	Q7SVF5 human immun
199	39	53.4	118	15	Q9ESP4	Q9ESP4 human immun
200	39	53.4	120	15	Q9IHU2	Q9IHU2 human immun
201	39	53.4	122	15	Q90DR1	Q90DR1 human immun
202	39	53.4	122	15	Q90DL1	Q90DL1 human immun
203	39	53.4	122	15	Q9YXR8	Q9YXR8 human immun
204	39	53.4	122	15	Q9IHX9	Q9IHX9 human immun
205	39	53.4	122	15	Q9QIV2	Q9QIV2 human immun
206	39	53.4	122	15	Q90DN9	Q90DN9 human immun
207	39	53.4	122	15	Q90DN4	Q90DN4 human immun
208	39	53.4	122	15	Q9EAA1	Q9EAA1 human immun
209	39	53.4	122	15	Q90DN3	Q90DN3 human immun
210	39	53.4	122	15	Q90DS7	Q90DS7 human immun
211	39	53.4	122	15	Q9WR01	Q9WR01 human immun
212	39	53.4	122	15	Q9EA83	Q9EA83 human immun
213	39	53.4	122	15	Q7ZJS5	Q7ZJS5 human immun
214	39	53.4	122	16	Q7VJ15	Q7VJ15 helicobacte
215	39	53.4	126	15	Q91HU1	Q91HU1 human immun
216	39	53.4	126	15	Q91HU3	Q91HU3 human immun
217	39	53.4	126	15	Q91HU6	Q91HU6 human immun
218	39	53.4	127	10	Q8H585	Q8H585 oryza sativ
219	39	53.4	130	15	Q90Q13	Q90Q13 human immun
220	39	53.4	130	15	Q91HV2	Q91HV2 human immun
221	39	53.4	130	15	Q9WR04	Q9WR04 human immun
222	39	53.4	130	15	Q91HU9	Q91HU9 human immun
223	39	53.4	131	15	Q9WR05	Q9WR05 human immun
224	39	53.4	131	15	Q91HU4	Q91HU4 human immun
225	39	53.4	131	15	Q91HU0	Q91HU0 human immun
226	39	53.4	132	15	Q91HV0	Q91HV0 human immun
227	39	53.4	133	15	Q8UR70	Q8UR70 human immun
228	39	53.4	133	15	Q90Q44	Q90Q44 human immun
229	39	53.4	133	15	Q90Q06	Q90Q06 human immun
230	39	53.4	133	15	Q90Q31	Q90Q31 human immun
231	39	53.4	133	15	Q8UR36	Q8UR36 human immun
232	39	53.4	133	15	Q8UR35	Q8UR35 human immun
233	39	53.4	133	16	Q9HT46	Q9HT46 pseudomonas
234	39	53.4	135	15	Q9DQ19	Q9DQ19 human immun
235	39	53.4	137	15	Q91HV5	Q91HV5 human immun

236	39	53.4	140	15	Q9IEB9	Q9ieb9 human immun
237	39	53.4	144	15	Q7ZCE6	Q7zce6 human immun
238	39	53.4	144	15	Q7ZC77	Q7zc77 human immun
239	39	53.4	145	15	Q7ZC44	Q7zc44 human immun
240	39	53.4	148	10	Q8S673	Q8s673 oryza sativ
241	39	53.4	153	13	Q8J3R1	Q8j3r1 human immun
242	39	53.4	153	15	Q7SM13	Q7sm13 human immun
243	39	53.4	155	15	Q8J3N5	Q8j3n5 human immun
244	39	53.4	155	15	Q8J3Q4	Q8j3q4 human immun
245	39	53.4	157	16	Q49991	Q49991 mycobacteri
246	39	53.4	158	15	Q8J3N6	Q8j3n6 human immun
247	39	53.4	158	16	Q8DLY4	Q8dly4 wigglewort
248	39	53.4	159	16	Q8KF27	Q8kf27 bacillus ha
249	39	53.4	161	16	Q07200	Q07200 mycobacteri
250	39	53.4	161	16	Q7TT30	Q7tt30 mycobacteri
251	39	53.4	165	2	Q93UE9	Q93ue9 uncultured
252	39	53.4	165	2	Q93UF3	Q93uf3 uncultured
253	39	53.4	165	15	Q9IE94	Q9ie94 human immun
254	39	53.4	172	15	Q9IEB3	Q9ieb3 human immun
255	39	53.4	174	15	Q9IEB1	Q9ieb1 human immun
256	39	53.4	174	15	Q9IEC0	Q9iec0 human immun
257	39	53.4	177	2	Q8RLI9	Q8rli9 gamma-prote
258	39	53.4	177	15	Q8JER7	Q8je7 human immun
259	39	53.4	180	15	Q8JAK4	Q8jak4 human immun
260	39	53.4	181	15	Q8JAK2	Q8jak2 human immun
261	39	53.4	183	15	Q9IEC1	Q9iec1 human immun
262	39	53.4	188	4	Q9XU00	Q9xu00 homo sapien
263	39	53.4	188	4	Q9BU41	Q9bu41 homo sapien
264	39	53.4	188	11	Q9D938	Q9d938 mus musculus
265	39	53.4	198	16	Q8PLY4	Q8ply4 xanthomonas
266	39	53.4	198	16	Q8PA94	Q8pa94 xanthomonas
267	39	53.4	200	10	Q9LTK5	Q9ltk5 arabidopsis
268	39	53.4	209	15	Q9IE66	Q9ie66 human immun
269	39	53.4	210	15	Q9IEB7	Q9ieb7 human immun
270	39	53.4	213	15	Q9IEC4	Q9iec4 human immun
271	39	53.4	213	15	Q9IEC3	Q9iec3 human immun
272	39	53.4	216	15	Q9IEC5	Q9iec5 human immun
273	39	53.4	216	15	Q9IEC7	Q9iec7 human immun
274	39	53.4	216	15	Q9IEA5	Q9iea5 human immun
275	39	53.4	216	15	Q9IE98	Q9ie98 human immun
276	39	53.4	218	15	Q9IEA4	Q9iea4 human immun
277	39	53.4	218	15	Q9IE95	Q9ie95 human immun
278	39	53.4	219	15	Q9IEB6	Q9ieb6 human immun
279	39	53.4	219	15	Q9IEC8	Q9iec8 human immun
280	39	53.4	220	15	Q9IE97	Q9ie97 human immun
281	39	53.4	222	12	Q89245	Q89245 woodchuck h
282	39	53.4	222	12	Q83760	Q83760 woodchuck h
283	39	53.4	224	15	Q9IEA8	Q9iea8 human immun
284	39	53.4	225	15	Q9IEA0	Q9iea0 human immun
285	39	53.4	226	15	Q9IE93	Q9ie93 human immun
286	39	53.4	227	15	Q9IE99	Q9ie99 human immun
287	39	53.4	230	15	Q9IEA9	Q9iea9 human immun
288	39	53.4	232	15	Q9IE61	Q9ie61 human immun
289	39	53.4	232	15	Q9IE64	Q9ie64 human immun
290	39	53.4	234	15	Q9IEC2	Q9iec2 human immun
291	39	53.4	234	15	Q9IEA6	Q9iea6 human immun
292	39	53.4	235	15	Q9IE54	Q9ie54 human immun
293	39	53.4	238	15	Q9DIX1	Q9dix1 human immun
294	39	53.4	240	15	Q9IE32	Q9ie32 human immun
295	39	53.4	242	15	Q9IE31	Q9ie31 human immun
296	39	53.4	242	15	Q9IE30	Q9ie30 human immun
297	39	53.4	243	15	Q9IE55	Q9ie55 human immun
298	39	53.4	264	16	Q8EX01	Q8ex01 mycoplasma
299	39	53.4	266	16	Q8YC37	Q8yc37 brucella me
300	39	53.4	266	16	Q8FW88	Q8fw88 brucella su

ALIGNMENTS

RESULT 1
Q87G63
ID Q87G63

PRELIMINARY; PRT; 328 AA.

AC Q87G63;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Conserved hypothetical protein.
GN VPA1454.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIWD_2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739;
RA Makino K., Kishima K., Nakano M., Yokoyama K., Uda T., Tagomori K.,
Iijima Y., Nishimura K., Nakano M., Kubota Y., Kimura S.,
Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RA "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
distinct from that of V. cholerae.";
RL Lancet 361:743-749(2003).
DR EMBL; AF005089; BAC62797.1; -.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PR00453; VWFADOMAIN.
DR PROSITE; PS50234; VWF_A; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 328 AA; 7036 MW; DESF2BF556D904E5 CRC64;

Query Match 65.1%; Score 47.5; DB 16; Length 328;
Best Local Similarity 58.3%; Pred. No. 43;
Matches 7; Conservative 1; Mismatches 1; Indels 3; Gaps 1;

QY 1 EWTNI---WWWA 9
Db :||| ||||
6 QWLNIEFWWWA 17

RESULT 2
Q85FP6 PRELIMINARY; PRT; 340 AA.
ID Q85FP6;
AC Q85FP6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ORF340.
OS Cyanidioschyzon merolae (Red alga).
OC Eukaryota; Rhodophyta; Bangiophyceae; Cyanidioschyzon.
OX NCBI_TaxID=45157;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=10D;
RX MEDLINE=22639682; PubMed=12755171;
RA Ohta N., Matsuzaki M., Misumi O., Miyagishima S., Nozaki H.,
Tanaka K., Shin-i T., Kohara Y., Kuroiwa T.;
RA "Complete Sequence and Analysis of the Plastid Genome of the
RT Unicellular Red Alga Cyanidioschyzon merolae.";
RL DNA Res. 10:67-77(2003).
DR EMBL; AB002583; BAC76299.1; -.
DR GO; GO:0009507; C:chloroplast; IEA.
KW Chloroplast.
SQ SEQUENCE 340 AA; 40768 MW; 3F05B9895B3891C5 CRC64;

Query Match 64.4%; Score 47; DB 8; Length 340;
Best Local Similarity 71.4%; Pred. No. 52;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 WTNINWW 8
Db :||| ||||
93 WKQIWWW 99

RESULT 3


```
Q94RJ3
ID Q94RJ3 PRELIMINARY; PRT; 458 AA.
AC Q94RJ3
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE NADH subunit 4 (EC 1.6.5.3) (NADH-ubiquinone oxidoreductase chain 4)
DE (Fragment).
DE NADH4
GN Chimera monstrosa (Rabbit fish).
OS Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Holocephala; Chimaeiriformes; Chimaeiridae; Chimaeira.
OX NCBI_TaxID=7871;
RN [1]
SEQUENCE FROM N.A.
RA Arnason U., Gullberg A., Janke A.;
RC "Molecular phylogenetics of gnathostomous (jawed) fishes: Old bones,
RT new cartilage.";
RL Zool. Scr. 30:249-255(2001).
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
DR EMBL; A310140; CAC84208.1; -.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006120; F:mitochondrial electron transport, NADH to u. . .; IEA.
DR InterPro; IPR003918; NADH_oxred4.
DR InterPro; IPR001750; Oxidored_g1.
DR InterPro; IPR000260; Oxidored_g5_N.
DR Pfam; PF00361; oxidored_g1; 1.
DR PRINTS; PR01437; NUOXDETASE4.
KW NAD; Oxidoreductase; Ubiquinone; Mitochondrion.
FT NON_TER 458 458
SQ SEQUENCE 458 AA; 51775 MW; A9DFFFEA9C1525EA CRC64;

Query Match 63.0%; Score 4%; DB 8; Length 458;
Best Local Similarity 75.0%; Pred. No. 94;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNIW 8
DE 189 EWANILW 196

RESULT 4
Q90DP7 PRELIMINARY; PRT; 122 AA.
AC Q90DP7
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope glycoprotein (fragment).
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
SEQUENCE FROM N.A.
RA Hu D.J., Baggs J., Downing R.G., Pieniazek D., Dorn J., Fridlund C.,
RA Biryahwahwo B., Sempala S.D.K., Rayfield M.A., Dondero T.J., Lal R.;
RT "Similar distribution and continued predominance of HIV-1 subtypes A
RT and D infections in Uganda."
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF307684; AAL08745.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env_GP41.
DR Pfam; PF00517; GP41; 1.
KW Transmembrane.
FT NON_TER 1 1
```

```
FT NON_TER 122 122
SQ SEQUENCE 122 AA; 14783 MW; BD6F1DE521F2BCBE CRC64;

Query Match 61.6%; Score 45; DB 15; Length 122;
Best Local Similarity 62.5%; Pred. No. 37;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNIW 8
DE 105 KWTNLTW 112

RESULT 5
Q8ZQ42 PRELIMINARY; PRT; 168 AA.
AC Q8ZQ42
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Suppression of copper sensitivity: lipoprotein modification in lgt
DE mutants of E. coli.
GN SCSD OR STM1116.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
SEQUENCE FROM N.A.
RA STRAIN=172 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
DR EMBL; A5008748; AAL20348.1; -.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR006662; Thioired.
DR InterPro; IPR006663; Thioiredox_dom2.
DR Pfam; PF00085; thioired; 1.
KW Lipoprotein; Complete proteome.
SQ SEQUENCE 168 AA; 18549 MW; 7F498959A5683149 CRC64;

Query Match 61.6%; Score 45; DB 16; Length 168;
Best Local Similarity 46.7%; Pred. No. 50;
Matches 7; Conservative 1; Mismatches 1; Indels 6; Gaps 1;

QY 2 WTNIW-----WWAK 10
DE 152 WTSYGMKRLRWAK 166

RESULT 6
Q8Z7P2 PRELIMINARY; PRT; 168 AA.
AC Q8Z7P2
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Secreted protein, suppressor for copper-sensitivity D (Secreted
DE copper-sensitivity suppressor D).
GN SCSD OR STV1152 OR T1804.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
SEQUENCE FROM N.A.
RA STRAIN=CT18;
```

```
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churche C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Conerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krcgh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Sissons M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT *Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.*;
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Playhew G.F., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT *Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.*;
RL J. Bacteriol. 185:2330-2337(2003).
DR EMBL; AL627269; CA008241.1; -.
DR EMBL; AE016940; AA069426.1; -.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0096118; P:electron transport; IEA.
DR InterPro; IPR006662; Thiorod.
DR InterPro; IPR006663; Thiorodex_dom2.
DR Pfam; PF00085; Thiorod; 1.
KW Complete proteome.
SQ SEQUENCE 168 AA; 18423 MW; 954D9B73A84A2D81 CRC64;

Query Match 61.6%; Score 45; DB 16; Length 168;
Best Local Similarity 46.7%; Pred. No. 50;
Matches 7; Conservative 1; Mismatches 1; Indels 6; Gaps 1;

QY 2 WTNIW-----WWAK 10
Db 152 WTSYWGKMLRLWAK 166
|||:|
|||:|

RESULT 7
Q8AU06 PRELIMINARY; PRT; 211 AA.
ID Q8AU06
AC Q8AU06;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=03LTS;
RA Nanteza M.B., Virrell D.L., Kintu P., Kaleebu P., Biryahwaho B.,
RA Morgan D., Whitworth J.;
RT *Correlation of tat, nef and the membrane proximal cytoplasmic domain
RT of gp41 to HIV-1 disease progression in Uganda.*;
RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF425920; AA031576.1; -.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env_GP41.
DR Pfam; PF00517; GP41; 1.
FT NCN TER 1
FT NCN TER 211
FT NCN TER 211
SQ SEQUENCE 211 AA; 24919 MW; DCC8B3A06DF5003 CRC64;

Query Match 61.6%; Score 45; DB 15; Length 211;
Best Local Similarity 62.5%; Pred. No. 62;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 EWTNIWW 8
Db 58 KWTNLWTW 65
|||:|
|||:|

RESULT 8
Q64901 PRELIMINARY; PRT; 222 AA.
ID Q64901
AC Q64901;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Small envelope protein (Major surface antigen).
GN S.
OS Arctic ground squirrel hepatitis B virus.
OC Viruses; Retrovirdae; Hepadnaviridae; Orthohepadnavirus.
OX NCBI_TaxID=41952;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96256727; PubMed=8676441;
RA Testut P., Renard C.A., Terradillos O., Vitvitski-Trepa L., Tekala F.,
RA Degott C., Blake J., Boyer B., Buendia M.A.;
RT *A new hepadnavirus endemic in arctic ground squirrels in Alaska.*;
RL J. Virol. 70:4210-4219(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Testut P., Buendia M.A.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U29144; AA08035.1; -.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0016032; P:viral life cycle; IEA.
DR InterPro; IPR000349; Hepvir_surfa.
DR Pfam; PF00695; VMSA; 1.
KW Antigen; Envelope protein.
SQ SEQUENCE 222 AA; 25410 MW; C08FA95C667B7489 CRC64;

Query Match 61.6%; Score 45; DB 12; Length 222;
Best Local Similarity 46.7%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 2 WTNI-----WW 8
Db 20 WTNLTIAQSLDWW 34
|||:|
|||:|

RESULT 9
Q64900 PRELIMINARY; PRT; 282 AA.
ID Q64900
AC Q64900;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Middle envelope protein (Major surface antigen).
GN PRES2/S.
OS Arctic ground squirrel hepatitis B virus.
OC Viruses; Retrovirdae; Hepadnaviridae; Orthohepadnavirus.
OX NCBI_TaxID=41952;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96256727; PubMed=8676441;
RA Testut P., Renard C.A., Terradillos O., Vitvitski-Trepa L., Tekala F.,
RA Degott C., Blake J., Boyer B., Buendia M.A.;
RT *A new hepadnavirus endemic in arctic ground squirrels in Alaska.*;
RL J. Virol. 70:4210-4219(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Testut P., Buendia M.A.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U29144; AA08034.1; -.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0016032; P:viral life cycle; IEA.
DR InterPro; IPR000349; Hepvir_surfa.
DR Pfam; PF00695; VMSA; 1.
```

```

KW Antigen; Envelope protein.
SQ SEQUENCE 282 AA; 31803 MW; B0DF588638009D02 CRC64;

Query Match 61.6%; Score 45; DB 12; Length 282;
Best Local Similarity 46.7%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Cy 2 WTNI-----WWW 8
    |||||
Db 80 WTNIITTAQSLDWW 94

RESULT 10
ID Q64899 PRELIMINARY; PRT; 427 AA.
AC Q64899;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Large envelope protein (Major surface antigen).
CN PRES1/PRES2/S.
OS Arctic ground squirrel hepatitis B virus.
OC Viruses; Retroviruses; Hepadnaviridae; Orthohepadnavirus.
CX NCBI_TaxID=41952;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96256727; PubMed=8676441;
RA Testut P., Renard C.A., Terradillos O., Vitvitski-Trepo L., Tekala F.,
RA Degott C., Blake J., Boyer B., Buendia M.A.;
RT "A new hepadnavirus endemic in arctic ground squirrels in Alaska.";
RL J. Virol. 70:4210-4219(1996).
[2]
RP SEQUENCE FROM N.A.
RA Testut P., Buendia M.A.;
EL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U29144; AB08033.1; -.
DR GO; GO:0019031; C:viral envelope; IEA.
DR InterPro; IPR000349; Hepvir_surfa.
DR Pfam; PF00695; vMSA; 1.
KW Antigen; Envelope protein.
SQ SEQUENCE 427 AA; 48270 MW; A2AE1EB5707DF940 CRC64;

Query Match 61.6%; Score 45; DB 12; Length 427;
Best Local Similarity 46.7%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

Cy 2 WTNI-----WWW 8
    |||||
Db 225 WTNIITTAQSLDWW 239

RESULT 11
ID Q90QM3 PRELIMINARY; PRT; 860 AA.
AC Q90QM3;
DT 01-DEC-2001 (TREMBLrel. 13, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroviruses; Retroviridae; Lentivirus.
CX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=94CY049-93e;
RA Gao F., Vidal N., Li Y., Trask S.A., Chen Y., Kostrikis L.G., Ho D.D.,
RA Kim J., Cioe K., Oh M.-D., Salminen M., Robertson D.L., Shaw G.M.,
RA Hahn B.H., Peeters M.;
RT "Evidence for two distinct sub-types within the HIV-1 subtype A
RT radiation.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

```

```

DR EMBL; AF286254; AAK82678.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env GP41.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.
FT NON_TER 1
SQ SEQUENCE 860 AA; 97493 MW; 927D148A584208B1 CRC64;

Query Match 61.6%; Score 45; DB 15; Length 860;
Best Local Similarity 62.5%; Pred. No. 2.3e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Cy 1 EWTNIWWW 8
    :|||:|
Db 662 KWTNLWTW 669

RESULT 12
ID Q90VI2 PRELIMINARY; PRT; 868 AA.
AC Q90VI2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Envelope glycoprotein.
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroviruses; Retroviridae; Lentivirus.
CX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=FIN91121;
RC MEDLINE=20506840; PubMed=11055649;
RA Liitsola K., Holmstrom P., Laukkanen T., Brummer-Korvenkontio H.,
RA Leinikki P., Salminen M.O.;
RT "Analysis of HIV-1 genetic subtypes in Finland reveals good
RT correlation between molecular and epidemiological data.";
RL Scand. J. Infect. Dis. 32:475-480(2000).
DR EMBL; AF219261; AAKS3072.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env GP41.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.
SQ SEQUENCE 868 AA; 97966 MW; 9B1E1AADB49213EA CRC64;

Query Match 61.6%; Score 45; DB 15; Length 868;
Best Local Similarity 62.5%; Pred. No. 2.3e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Cy 1 EWTNIWWW 8
    :|||:|
Db 670 KWTNLWTW 677

RESULT 13
ID Q9YV03 PRELIMINARY; PRT; 868 AA.
AC Q9YV03;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Envelope glycoprotein.
GN ENV.

```

```

CS Human immunodeficiency virus 1.
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=SE7535;
RX MEDLINE=9441797; PubMed=10513639;
RA Carr J.K., Laukkanen T., Salminen M.O., Albert J., Alaeus A., Kim B.,
RA Sanders-Buell E., Birk D.L., McCutchan F.E.;
RT "Characterization of subtype A HIV-1 from Africa by full genome
RT sequencing.";
RL AIDS 13:1819-1826(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SE7535;
RA Carr J.K., Kim B., Sanders-Buell E., Salminen M.O., Alaeus A.,
RA Albert J.A., Birk D.L., McCutchan F.E.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF069671; AAC69304.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env_GP41.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.
SQ SEQUENCE 868 AA; 98393 MW; 72EE54428811EDE2 CRC64;

Query Match 61.6%; Score 45; DB 15; Length 868;
Best Local Similarity 62.5%; Pred. No. 23+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNLMWW 8
Db 670 KWTNLMTW 677

RESULT 14
Q85FZ6 PRELIMINARY; PRT; 47 AA.
AC Q85FZ6;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE GRP47.
GN YCP33.
OS Cyanidioschyzon merolae (Red alga).
OG Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Cyanidioschyzon.
OX NCBI_TaxID=45157;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=10D;
RX MEDLINE=22639682; PubMed=12755171;
RA Ohta N., Matsuzaki M., Mishima O., Miyagishima S., Nozaki H.,
RA Tanaka K., Shin-i T., Kohara Y., Kuroiwa T.;
RT "Complete Sequence and Analysis of the Plastid Genome of the
RT Unicellular Red Alga Cyanidioschyzon merolae.";
RL DNA Res. 10:67-77(2003).
DR EMBL; AB002583; BAC76196.1; -.
DR GO; GO:0009507; C:chloroplast; IEA.
KW Chloroplast.
SQ SEQUENCE 47 AA; 5523 MW; 4E46903BCA0D3A13 CRC64;

Query Match 60.3%; Score 44; DB 8; Length 47;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNLMWW 8
Db 19 KWSPPFWW 26

```

```

RESULT 15
Q9YYZ6 PRELIMINARY; PRT; 117 AA.
AC Q9YYZ6;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE GP41 (Fragment).
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=184.574;
RA Brennan C.A., Lund J.K., Golden A., Yamaguchi J., Vallari A.S.,
RA Phillips J.F., Kataaha P.K., Jackson J.B., Devare S.G.;
RT "Serologic and Phylogenetic Characterization of HIV-1 Subtypes in
RT Uganda.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF006882; AAD01326.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env_GP41.
DR Pfam; PF00517; GP41; 1.
KW Transmembrane.
FT NON_TER 1
FT NON_TER 117
SQ SEQUENCE 117 AA; 14055 MW; 4FCEBE1B68CE208 CRC64;

Query Match 60.3%; Score 44; DB 15; Length 117;
Best Local Similarity 62.5%; Pred. No. 48;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNLMWW 8
Db 109 KWTNLMSW 116

RESULT 16
Q90DR7 PRELIMINARY; PRT; 122 AA.
AC Q90DR7;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=UG;
RA Hu D.J., Baggs J., Downing R.G., Pieniazek D., Dorn J., Fridlund C.,
RA Biryahwaho B., Sempala S.D.K., Rayfield M.A., Dondero T.J., Lal R.;
RT "Similar distribution and continued predominance of HIV-1 subtypes A
RT and D infections in Uganda.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF307664; AAL08725.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env_GP41.
DR Pfam; PF00517; GP41; 1.
KW Transmembrane.
FT NON_TER 1
FT NON_TER 122
SQ SEQUENCE 122 AA; 14693 MW; D20268AD4ADF8DEE CRC64;

```

```
Query Match 60.3%; Score 44; DB 15; Length 122;
Best Local Similarity 62.5%; Pred. No. 50;
Matches 5; Conservative 2; Mismatches 0; Gaps 0;

QY 1 EWTNIWW 8
DB 105 KWTNLMSW 112

RESULT 17
Q3IUM9 PRELIMINARY; PRT; 122 AA.
ID Q9IUM9
AC Q9IUM9
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retrovird viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=AR60;
RX MEDLINE=20346416; PubMed=10890362;
RA Masciotra S.; Livellara B.; Bellosso W.; Clara L.; Tanuri A.; Ramos A.;
RA Baggs J.; Lal R.; Pieniazek D.;
RT "Evidence for a high frequency of HIV-1 subtype F infections among
RT heterosexual population in Buenos Aires, Argentina.";
RL AIDS Res. Hum. Retroviruses 16:1007-1014 (2000).
DR EMBL; AF220703; AAF76822.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env_GP41.
DR Pfam; PF00517; GP41; 1.
KW Transmembrane.
FT NON TER 1
FT NON TER 122
SQ SEQUENCE 122 AA; 14679 MW; ABE4BB42IA98FFA5 CRC64;

Query Match 60.3%; Score 44; DB 15; Length 122;
Best Local Similarity 62.5%; Pred. No. 50;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNIWW 8
DB 105 KWTNLMSW 112

RESULT 18
C8JEQ8 PRELIMINARY; PRT; 174 AA.
ID Q8JEQ8
AC Q8JEQ8
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DE 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Envelope glycoprotein (Fragment).
OS Human immunodeficiency virus 1.
OC Viruses; Retrovird viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=JASN-0;
RX Poveda E.; Rodas B.; Toro C.; Martin-Carbonero L.; Soriano V.;
RT "Evolution of the env gene (gp41) in HIV-1 positive patients receiving
RT T-20, a fusion inhibitor.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF500092; AAM21678.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR003328; Env_GP41.

Query Match 60.3%; Score 44; DB 15; Length 324;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EWTNIWWAK 10
DB 268 EWSGAWWYEK 277

RESULT 20
Q24539 PRELIMINARY; PRT; 351 AA.
ID Q24539
AC Q24539
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE 39 kDa protein.
GN SU(F) OR CG17170.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]_
RP SEQUENCE FROM N.A.
RX MEDLINE=93170663; PubMed=8436295;
RA Mitchelson A.; Simoniellig M.; Williams C.; O'Hare K.;
```

"Homology with Saccharomyces cerevisiae RNA14 suggests that phenotypic suppression in Drosophila melanogaster by suppressor of forked occurs at the level of RNA stability.";

RT Genes Dev. 7:241-249 (1993).
 RL EMBL; X62679; CAA44552.1; -.
 DR EMBL; X62679; CAA44552.1; -.
 DR PIR; B46389; B46389.
 DR FLYBASE; FBgn0003559; su(f).
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0006396; P:RNA processing; IEA.
 DR InterPro; IPR003107; HAT.
 DR InterPro; IPR008847; SuF.
 DR InterPro; IPR008941; TPR-like.
 DR Pfam; PF05843; SuF; 1.
 DR SMART; SM00386; HAT; 4.
 SQ SEQUENCE 351 AA; 41525 MW; 038616A0FEEBF3B CRC64;

Query Match 60.3%; Score 44; DB 5; Length 351;
 Best Local Similarity 58.3%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 3; Indels 2; Gaps 1;

QY 1 EWTN--IWWAK 10
 |||
 Db 334 EWVMAFAWWAK 345

RESULT 21

Q8MZ12 PRELIMINARY; PRT; 356 AA.
 AC Q8MZ12;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE GH16848P (Fragment).
 SUF) OR CG17170.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Friese E.,
 RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Saragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celniker S.;
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY102664; AAM27493.1; -.
 DR FLYBASE; FBgn0003559; su(f).
 DR InterPro; IPR008847; SuF.
 DR InterPro; IPR008941; TPR-like.
 DR Pfam; PF05843; SuF; 1.
 FT NON TER 356 356
 SQ SEQUENCE 356 AA; 42166 MW; 8DAF19BF81038616 CRC64;

Query Match 60.3%; Score 44; DB 5; Length 356;
 Best Local Similarity 58.3%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 3; Indels 2; Gaps 1;

QY 1 EWTN--IWWAK 10
 |||
 Db 334 EWVMAFAWWAK 345

RESULT 22

Q8Q7H3 PRELIMINARY; PRT; 876 AA.
 AC Q8Q7H3;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)

DB Envelope glycoprotein.
 GN ENV.
 OS Human immunodeficiency virus 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=8CMAJ323;
 RX MEDLINE=21849375; PubMed=11860674;
 RA Yamaguchi J., Vallari A.S., Swanson P., Bodelle P., Kaptue L.,
 RA Ngamsoo C., Zekeng L., Gurtler L.G., Devare S.G., Brennan C.A.;
 RT "Evaluation of HIV Type 1 Group O Isolates: Identification of Five
 RT Phylogenetic Clusters";
 RL AIDS Res. Hum. Retroviruses 18:269-282 (2002).
 DR EMBL; AF383248; AAL98870.1; -.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR000328; Env GP41.
 DR InterPro; IPR000777; GP120.
 DR Pfam; PF00516; GP120; 1.
 DR Pfam; PF00517; GP41; 1.
 KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.
 SQ SEQUENCE 876 AA; 98131 MW; D313E4314FEE17A CRC64;

Query Match 60.3%; Score 44; DB 15; Length 876;
 Best Local Similarity 75.0%; Pred. No. 3.2e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EMTN1WWW 8
 |||
 Db 678 EMTN1WWW 685

RESULT 23

Q8KU32 PRELIMINARY; PRT; 423 AA.
 AC Q8KU32;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Putative phospholipase C.
 GN PLCA.
 OS Legionella pneumophila.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
 OC Legionellaceae; Legionella.
 OX NCBI_TaxID=446;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=130b;
 RX MEDLINE=22095963; PubMed=12101309;
 RA Aragon V., Rossier O., Cianciotto N.P.;
 RT "Legionella pneumophila genes that encode lipase and phospholipase C
 RT activities";
 RL Microbiology 148:2223-2231 (2002).
 DR EMBL; AF454865; AAM73854.1; -.
 SQ SEQUENCE 423 AA; 48028 MW; A613B1F6853C54F9 CRC64;

Query Match 59.6%; Score 43.5; DB 2; Length 423;
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 2 WTN1WW-NA 9
 |||
 Db 378 WTN1WW-NA 386

RESULT 24

Q69908 PRELIMINARY; PRT; 49 AA.
 AC Q69908;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE GP41 (Fragment).
 GN ENV.

OS Human immunodeficiency virus 1.
 OS Viruses; Retroviral viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=94211861; PubMed=7512731;
 RA Conley A.J., Kessler J.A., Boots L.J., Tung J.S., Arnold B.A.,

RA Keller P.M., Shaw A.R., Enini E.A.;
 RT "Neutralization of divergent human immunodeficiency virus type 1
 variants and primary isolates by IAM-4-2F5, an anti-gp41 human

RT monoclonal antibody.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:3348-3352(1994).

DR EMBL; U06738; AAA19151.1; -;
 DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.

DR InterPro; IPR000328; Env_GP41.
 DR Pfam; PF00517; GP41; 1.

KW Transmembrane.
 FT NON_TER 1 1
 FT NON_TER 49 49

SQ SEQUENCE 49 AA; 6283 MW; 74CD75339B92C172 CRC64;
 Query Match 58.9%; Score 43; DB 15; Length 49;
 Best Local Similarity 62.5%; Pred. No. 29;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EWTNIWNW 8
 DB 41 KWTNLWNW 48

RESULT 25
 ID Q9LQNO PRELIMINARY; PRT; 95 AA.

AC Q9LQNO;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE F5D14.5.
 DE F5D14.5.

GN Arabidopsis thaliana (Mouse-ear cress).
 OS Arabidopsi

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Lia S.X., Chan A., Yu G., Lee J.M., Lenz C., Pham P., Sakano H.,

RA Toriumi M., Vysotskaya V.S., Chan C., Chioi J., Choi E., Chung M.,
 RA Gonzalez A., Hwang B., Liu A., Vaysberg M., Altati H., Brooks S.,

RA Buehler E., Chao Q., Conn L., Conway A.B., Hansen N.F.,
 RA Johnson-Hopson C., Khan S., Kim C., Lam B., Miranda M., Nguyen M.,

RA Palm C.J., Shinn P., Southwick A., Davis R.W., Ecker J.R.,
 RA Federspiel N.A.; Theologis A.;

RT "The sequence of SAC F5D14 from Arabidopsis thaliana chromosome 1.";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;
 RA Theologis A.;

RA Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC007767; AAFB1325.1; -;

DR PIR; E86447; E86447.
 SQ SEQUENCE 95 AA; 10624 MW; 0C33985771E8B54E CRC64;

Query Match 58.9%; Score 43; DB 10; Length 95;
 Best Local Similarity 62.5%; Pred. No. 54;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 EWTNIWNW 8
 DB 75 EWSWWWWW 82

RESULT 26
 ID Q90DK4 PRELIMINARY; PRT; 122 AA.

AC Q90DK4;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Envelope glycoprotein (Fragment).
 GN ENV.

OS Human immunodeficiency virus 1.
 OS Viruses; Retroviral viruses; Retroviridae; Lentivirus.

OX NCBI_TaxID=11676;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=UG;

RA Hu D.J., Baggs J., Downing R.G., Pieniazek D., Dorn J., Fridlund C.,
 RA Biryahwaho B., Sempala S.D.K., Rayfield M.A., Dondoro T.J., Lal R.;

RT "Similar distribution and continued predominance of HIV-1 subtypes A
 and D infections in Uganda.";

RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF307727; AAL08788.1; -;

DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR000328; Env_GP41.

DR Pfam; PF00517; GP41; 1.
 KW Transmembrane.

FT NON_TER 1 1
 FT NON_TER 122 122

SQ SEQUENCE 122 AA; 14735 MW; 358D3FFEFB3C525 CRC64;
 Query Match 58.9%; Score 43; DB 15; Length 122;
 Best Local Similarity 62.5%; Pred. No. 69;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EWTNIWNW 8
 DB 105 KWTNLWNW 112

RESULT 27
 ID Q90Q62 PRELIMINARY; PRT; 133 AA.

AC Q90Q62;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Envelope glycoprotein (Fragment).
 GN ENV.

OS Human immunodeficiency virus 1.
 OS Viruses; Retroviral viruses; Retroviridae; Lentivirus.

OX NCBI_TaxID=11676;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=99ES-MOI469;

RX MEDLINE=21322034; PubMed=11429126;
 RA Ortiz M., Sanchez I., Gonzalez M.P., Leon M.I., Abeso N., Asumu E.,

RA Garcia-Saiz A.;
 RT "Molecular epidemiology of HIV type 1 subtypes in Equatorial Guinea.";

RL AIDS Res. Hum. Retroviruses 17:851-855(2001).
 DR EMBL; AF331079; AAK92290.1; -;

DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR InterPro; IPR000328; Env_GP41.

DR Pfam; PF00517; GP41; 1.

```

KW Transmembrane.
FT MON_TER 1
FT MON_TER 133
SQ SEQUENCE 133 AA; 15671 MW; D39665DE6DC18F35 CRC64;

Query Match 58.9%; Score 43; DB 15; Length 133;
Best Local Similarity 62.5%; Pred. No. 74;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNIWW 8
Db 122 KWTNLWW 129

RESULT 28
Q823W0 PRELIMINARY; PRT; 143 AA.
ID Q823W0
AC Q823W0;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DE 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Hypothetical protein.
GN CCA00295.
OS Chlamydothila caviae.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydothila.
OX NCBI_TaxID=83557;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=GPIC;
RA Read T.D., Myers G.S.A., Brunham R.C., Nelson W.C., Paulsen I.T.,
RA Heidelberg J., Holtzapflee E., Khouri H., Federova N.B., Carty H.A.,
RA Umayam L.A., Haft D.H., Peterson J., Beanan M.J., White O.,
RA Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G., Bavol P.M.,
RA Fraser C.M.;
RT "Genome sequence of Chlamydothila caviae (Chlamydia psittaci GPIC):
RT examining the role of niche-specific genes in the evolution of the
RT Chlamydiaceae."
RL Nucleic Acids Res. 31:2134-2147(2003).
DR EMBL; AE016995; AAP05044.1; -.
DR TIGR; CCA00295; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 143 AA; 16513 MW; 8864CE2CA138CF63 CRC64;

Query Match 58.9%; Score 43; DB 16; Length 143;
Best Local Similarity 62.5%; Pred. No. 80;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WTNIIWWA 9
Db 117 WSNLWWA 124

RESULT 29
Q82MM4 PRELIMINARY; PRT; 392 AA.
ID Q82MM4
AC Q82MM4;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DE 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Putative glutamate dehydrogenase.
GN ROCG OR SAV1636.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
EX MEDLINE=2-477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose X., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;

```

```

RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites."
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis."
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AF005027; BAC69347.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:004420; F:hydroxymethylglutaryl-CoA reductase (NADPH) . . . ; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006520; P:amino acid metabolism; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR InterPro; IPR006095; GLFV dehydrog.
DR InterPro; IPR002202; HMG-CoA red.
DR InterPro; IPR001005; Myb DNA binding.
DR InterPro; IPR000205; NAD_BS.
DR Pfam; PF00208; GLFV dehydrog; 1.
DR PROSITE; PS00074; GLFV DEHYDROGENASE; 1.
DR PROSITE; PS00318; HMG COA REDUCTASE_2; 1.
DR PROSITE; PS00037; MYB_1; 1.
KW Complete proteome.
SQ SEQUENCE 392 AA; 41874 MW; FC7149EACE86E7F4 CRC64;

Query Match 58.9%; Score 43; DB 16; Length 392;
Best Local Similarity 83.3%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TNIWW 8
Db 324 TNANWW 329

RESULT 30
Q81V01 PRELIMINARY; PRT; 605 AA.
ID Q81V01
AC Q81V01;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Similar to protocadherin gamma subfamily C, 4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Brain;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC042650; AAH42650.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR Pfam; PF00028; cadherin; 3.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM0112; CA; 4.
DR PROSITE; PS00232; CADHERIN_1; 4.
DR PROSITE; PS00268; CADHERIN_2; 4.
SQ SEQUENCE 605 AA; 67077 MW; 11CF6634133F4C9B CRC64;

Query Match 58.9%; Score 43; DB 4; Length 605;
Best Local Similarity 75.0%; Pred. No. 3.1e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```



```
QY 2 WTNINWWA 9
   |||||
Db 8 WTEINWWA 15

RESULT 31
Q90VE9          PRELIMINARY;      PRT;      621 AA.
AC Q90VE9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope glycoprotein.
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FIN9379;
RX MEDLINE=20506840; PubMed=11055649;
RA Liitsola K., Holmstrom P., Laukkanen T., Brummer-Korvenkontio H.,
RA Leinikki P., Salminen M.O.;
RT "Analysis of HIV-1 genetic subtypes in Finland reveals good
RT correlation between molecular and epidemiological data.";
RL Scard. J. Infect. Dis. 32:475-480 (2000).
DR EMBL: AF219273; AAK53126.1; -.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR GO: GO:0019028; C: viral capsid; IEA.
DR GO: GO:0019031; C: viral envelope; IEA.
DR GO: GO:0005198; F: structural molecule activity; IEA.
DR InterPro: IPR000328; Env GP41.
DR EMBL: AF219273; GPl20.
DR Pfam: PF00516; GPl20; 1.
DR Pfam: PF00517; GPl20; 1.
KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.
SQ SEQUENCE 621 AA; 69497 MW; 2A7CEAD897AE737F CRC64;

Query Match      58.9%; Score 43; DB 15; Length 621;
Best Local Similarity 71.4%; Pred. No. 3.2e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 WTNINWW 8
   |||||
Db 10 WTNLW 16

RESULT 32
Q9WMU7          PRELIMINARY;      PRT;      804 AA.
AC Q9WMU7;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope (Gpl60) (Fragment).
GN ENV OR GPl60.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ZR36;
RX MEDLINE=95294894; PubMed=10364493;
RA Triques K., Bourgeois A., Saragosti S., Vidal N., Mpoudi-Etong E.,
RA Ntilambi N., Apere C., Ekwilanga M., Delaporte E., Peeters M.;
RT "High diversity of HIV-1 subtype F strains in Central Africa.";
RL Virology 259:99-109 (1999).
DR EMBL: AJ237809; CAB44305.1; -.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR GO: GO:0019028; C: viral capsid; IEA.
DR GO: GO:0019031; C: viral envelope; IEA.
DR GO: GO:0005198; F: structural molecule activity; IEA.

QY InterPro: IPR000328; Env GP41.
DR InterPro: IPR000777; GPl20.
DR Pfam: PF00516; GPl20; 1.
DR Pfam: PF00517; GPl20; 1.
KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.
SQ SEQUENCE 804 AA; 91106 MW; CE300CADC83D222F CRC64;

Query Match      58.9%; Score 43; DB 15; Length 804;
Best Local Similarity 62.5%; Pred. No. 4.1e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNINWW 8
   :|||:|
Db 623 KWTNLDNWW 630

RESULT 33
Q9WC60          PRELIMINARY;      PRT;      848 AA.
AC Q9WC60;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope glycoprotein.
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE9280;
RX MEDLINE=99160125; PubMed=10052760;
RA Laukkanen T., Albert J., Liitsola K., Green S.D., Carr J.K.,
RA Leitner T., McCutchan F.E., Salminen M.O.;
RT "Virtually full-length sequences of HIV type 1 subtype J reference
RT strains.";
RL AIDS Res. Hum. Retroviruses 15:293-297 (1999).
DR EMBL: AF082394; AAD17761.1; -.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR GO: GO:0019028; C: viral capsid; IEA.
DR GO: GO:0019031; C: viral envelope; IEA.
DR GO: GO:0005198; F: structural molecule activity; IEA.
DR InterPro: IPR000328; Env GP41.
DR InterPro: IPR000777; GPl20.
DR Pfam: PF00516; GPl20; 1.
DR Pfam: PF00517; GPl20; 1.
KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.
SQ SEQUENCE 848 AA; 95486 MW; 1FF5F2385E98E36E CRC64;

Query Match      58.9%; Score 43; DB 15; Length 848;
Best Local Similarity 62.5%; Pred. No. 4.3e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNINWW 8
   :|||:|
Db 657 KWTNLDNWW 664

RESULT 34
Q9WC69          PRELIMINARY;      PRT;      850 AA.
AC Q9WC69;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope glycoprotein.
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=SE9173;
RX MEDLINE=99160125; PubMed=10052760;
RA Laakkonen T., Albert J., Lilisola K., Green S.D., Carr J.K.,
RA Leitner T., McCutchan F.E., Salminen M.O.;
RT "Virtually full-length sequences of HIV type 1 subtype J reference
RT strains.";
RL AIDS Res. Hum. Retroviruses 15:293-297 (1999).
DR ENBL; AF082395; AAD1768.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env GP41.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.
SQ SEQUENCE 850 AA; 95856 MW; 738E4F5492C877C9 CRC64;

Query Match 58.9%; Score 43; DB 15; Length 850;
Best Local Similarity 62.5%; Pred. No. 4.3e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EWTNINWW 8
Db 659 KWTNINWW 666
:|||||

RESULT 35
O41564 PRELIMINARY; PRT; 856 AA.
AC O41564
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JUN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope glycoprotein.
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroviridae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C26;
RX MEDLINE=98105804; PubMed=9445059;
RA Connor R.I., Korber B.T.M., Graham B.S., Hahn B.H., Ho D.D.,
RA Walker B.D., Neumann A.U., Vermund S.H., Mestecky J., Jackson S.,
RA Fenamore E., Cao Y., Gao F., Kalams S., Kunstman K.J., McDonald D.,
RA McWilliams N., Trkola A., Moore J.P., Wolinsky S.M.;
RT "Immunological and virological analyses of persons infected by human
RT immunodeficiency virus type 1 while participating in trials of
RT recombinant gp120 subunit vaccines.";
RL J. Virol. 72:1552-1576 (1998).
DR ENBL; U84819; AAC58851.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env GP41.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.
SQ SEQUENCE 856 AA; 96843 MW; 71810C7AC032CA4D CRC64;

Query Match 58.9%; Score 43; DB 15; Length 856;
Best Local Similarity 62.5%; Pred. No. 4.3e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EWTNINWW 8
Db 665 KWTNINWW 672
:|||||

RESULT 36
O8Q870 PRELIMINARY; PRT; 868 AA.
AC O8Q870;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE GP160.
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroviridae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=KR3042;
RA Daniels R.S., Patel D., Xiang Z., Zheng N.N., Kang C.;
RT "Evidence for the spread of immune-escape HIV-1 subtype B in the
RT Korean population.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR ENBL; AJ417408; CAD10120.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env GP41.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
KW AIDS; Coat protein; Glycoprotein; Polyprotein; Transmembrane.
SQ SEQUENCE 868 AA; 98402 MW; 98B8AE4BFFCE585B CRC64;

Query Match 58.9%; Score 43; DB 15; Length 868;
Best Local Similarity 62.5%; Pred. No. 4.4e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EWTNINWW 8
Db 677 EWTSLMSW 684
:|||||

RESULT 37
O9Y5C3 PRELIMINARY; PRT; 871 AA.
AC O9Y5C3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Protocadherin gamma C4 short form protein.
GN PCDH-GAMMA-C4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99308636; PubMed=10380929;
RA Wu Q., Maniatis T.;
RT "A striking organization of a large family of human neural cadherin-
RT like cell adhesion genes.";
RL Cell 97:779-790 (1999).
DR ENBL; AF152525; AAD43785.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007153; F:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR Pfam; PF00028; cadherin; 5.
DR PRINTS; PR00205; CADHERIN.
DR SMART; SM00112; CA; 6.
DR PROSITE; PS00232; CADHERIN_1; 5.
DR PROSITE; PS50268; CADHERIN_2; 6.
```

```
KW Calcium; Calcium-binding; Cell adhesion; Glycoprotein.
SQ SEQUENCE 871 AA; 94476 MW; 7ECD96F291D15B11 CRC64;

Query Match 58.9%; Score 43; DB 4; Length 871;
Best Local Similarity 75.0%; Pred. No. 4.4e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 WTNWMA 9
   |||||
Db 8 WTEIWA 15

RESULT 38
ID Q9Y5F7 PRELIMINARY; PRT; 938 AA.
AC Q9Y5F7;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update);
DE Protocadherin gamma C4.
GN PCDH-GAMMA-C4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99308636; PubMed=10380929;
RA Wu Q., Maniatis T.;
RT "A striking organization of a large family of human neural cadherin-
RL like cell adhesion genes.";
RE Cell 97:779-790(1999)
DR ENBL; AFI52338; AAD43732.1; -.
DR Genew; HGNC:8717; PCDHGC4.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0005194; F:cell adhesion molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0007156; P:homophilic cell adhesion; IEA.
DR InterPro; IPR002126; Cadherin.
DR Pfam; PF00028; cadherin; 5.
DR PRINTS; PRO0205; CADHERIN.
DR SMART; SM00112; CA; 6.
DR PROSITE; PS00232; CADHERIN_1; 5.
DR PROSITE; PS0268; CADHERIN_2; 6.
KW Calcium; Calcium-binding; Cell adhesion; Glycoprotein.
SQ SEQUENCE 938 AA; 101213 MW; 99820B82A2F15CC CRC64;

Query Match 58.9%; Score 43; DB 4; Length 938;
Best Local Similarity 75.0%; Pred. No. 4.7e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 WTNWMA 9
   |||||
Db 8 WTEIWA 15

RESULT 39
ID Q9HAQ6 PRELIMINARY; PRT; 283 AA.
AC Q9HAQ6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created);
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ATP-binding cassette half-transporter (Fragment).
GN PR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RA Emadi-Konjin H.-P., Zhang H., Sun D., Schuetz J., Furuya K.N.;
RT "Isolation of a genomic clone containing the promoter region of the
RT human ATP-binding cassette (ABC) transporter HuPRP.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF308473; AAC33618.1; -.
DR GO; GO:0005524; P:ATP binding; IEA.
KW ATP-binding.
FT NON TER 283
SQ SEQUENCE 283 AA; 30876 MW; 3CD1395E7216BB47 CRC64;

Query Match 58.2%; Score 42.5; DB 4; Length 283;
Best Local Similarity 60.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 2 WTN-IWWAK 10
   |||||
Db 222 WNNPQWWAR 231

RESULT 40
ID Q9EX94 PRELIMINARY; PRT; 385 AA.
AC Q9EX94;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Phosphatidylcholine-hydrolyzing phospholipase C (EC 3.1.4.3).
GN PC-PLC.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=294;
RN [1]
RP SEQUENCE FROM N.A.
RA Preuss I., Kaiser I., Gehring U.;
RT "Molecular cloning of the phosphatidylcholine-hydrolyzing
RT phospholipase C from Pseudomonas fluorescens.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ304443; CAC18568.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004629; F:phospholipase C activity; IEA.
KW Hydrolase.
SQ SEQUENCE 385 AA; 42128 MW; 645090AED1A37D96 CRC64;

Query Match 58.2%; Score 42.5; DB 2; Length 385;
Best Local Similarity 66.7%; Pred. No. 2.4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 2 WTNWMA 9
   |||||
Db 362 WTDWGWMS 370

RESULT 41
ID Q9LCD3 PRELIMINARY; PRT; 508 AA.
AC Q9LCD3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Dipeptide transporter DppA homolog.
GN DppA.
OS Ochrobactrum anthropi.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Ochrobactrum.
OX NCBI_TaxID=529;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCRC SV3;
RX MEDLINE=20193627; PubMed=10727942;
RA Kameda H., Asano Y.;
RT "Gene cloning, nucleotide sequencing, and purification and
RT characterization of the D-stereospecific amino-acid amidase from
```

```
RT Ochrobactrium anthropi SV3."
RL Eur. J. Biochem. 267:2028-2035 (2000).
DR EMBL: AB028907; BAA94699.1; -.
DR GO: GO:0005215; P:transporter activity; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR000914; SBP_bac 5.
DR Pfam: PF00496; SBP_bac 5; 1.
DR PROSITE: PS01040; SBP_BACTERIAL_5; 1.
SQ SEQUENCE 508 AA; 55661 MW; 84E7579A526CA74E CRC64;

Query Match 58.2%; Score 42.5; DB 2; Length 508;
Best Local Similarity 50.0%; Pred. No. 3.1e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 3; Gaps 1;

QY 2 WTNI---WWAK 10
DB 491 WRNIATGWAWLR 502

RESULT 42
OS0329 PRELIMINARY; PRT; 104 AA.
AC OS0329;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Plasmid R388 genes.
OS Escherichia coli.
OG Plasmid R388.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K-12;
RX MEDLINE=97467347; PubMed=9325277;
RA Bolland S., Llosa M., de la Cruz F.;
RT "Genetic organization of the region involved in conjugative pilus
RT synthesis, export and assembly of the incW plasmid R388."
RL J. Biol. Chem. 272:25583-25590 (1997).
DR EMBL: X81123; CAA57023.1; -.
DR GO: GO:0046821; C:extrachromosomal DNA; IEA.
DR InterPro: IPR007792; VirB3.
DR Pfam: PF05101; VirB3; 1.
KW Plasmid.
SQ SEQUENCE 104 AA; 12234 MW; 652B59C4B1C5D4AD CRC64;

Query Match 57.5%; Score 42; DB 2; Length 104;
Best Local Similarity 83.3%; Pred. No. 81;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 NIWWWA 9
DB 46 SIWWWA 51

RESULT 43
OS2U3 PRELIMINARY; PRT; 108 AA.
AC OS2U3;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Nodule-specific glycine-rich protein 1A.
OS Medicago truncatula (Barrel medic).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Eukaryota; Tridiplantae; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids 1; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
OX NCBI_TaxID=3880;
RN [1]
RP SEQUENCE FROM N.A.
RA Kevei Z., Vinardell J.M., Kiss G.B., Kondorosi A., Kondorosi E.;
RT "Glycine-rich proteins encoded by a nodule-specific gene family are

RT implicated in different stages of symbiotic nodule development in
RT Medicago.";
RL Mol. Plant Microbe Interact. 0:0-0 (2002).
DR EMBL: AF498990; AAM18949.1; -.
SQ SEQUENCE 108 AA; 11271 MW; 93C9B8DACF14621B CRC64;

Query Match 57.5%; Score 42; DB 10; Length 108;
Best Local Similarity 44.4%; Pred. No. 83;
Matches 4; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 WTNIWWAK 10
DB 64 WGTSMWGR 72

RESULT 44
OS0DJ2 PRELIMINARY; PRT; 122 AA.
AC OS0DJ2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope glycoprotein (Fragment).
OS ENV.
GN Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UG;
RA Hu D.J., Baggs J., Downing R.G., Pieniazek D., Dorn J., Fridlund C.,
RA Biryahwaho B., Sempala S.D.K., Rayfield M.A., Dondero T.J., Lal R.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF327739; AAL08800.1; -.
DR GO: GO:0016021; C:integral to membrane; IEA.
DR GO: GO:0019031; C:viral envelope; IEA.
DR GO: GO:0005198; F:structural molecule activity; IEA.
DR InterPro: IPR000328; Env_GP41.
DR Pfam: PF00517; GP41; 1.
KW Transmembrane.
FT NON TER 1
FT NON TER 122
SQ SEQUENCE 122 AA; 14663 MW; AAC57DB097817188 CRC64;

Query Match 57.5%; Score 42; DB 15; Length 122;
Best Local Similarity 62.5%; Pred. No. 94;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNIWW 8
DB 105 EWANLWN 112

RESULT 45
OS1JK9 PRELIMINARY; PRT; 122 AA.
AC OS1JK9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope glycoprotein (Fragment).
OS ENV.
GN Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AR54;
RX MEDLINE=20346416; PubMed=10890362;
RA Masciotra S., Livellara B., Belloso W., Clara L., Tanuri A., Ramos A.,
RA Baggs J., Lal R., Pieniazek D.;
RT "Evidence for a high frequency of HIV-1 subtype F infections among
RT heterosexual population in Buenos Aires, Argentina.";
```

```
RL AIDS Res. Hum. Retroviruses 16:1007-1014 (2000).
DR EMBL: AF220273; AA576842.1; -.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR GO: GO:0019031; C: viral envelope; IEA.
DR GO: GO:0005198; F: structural molecule activity; IEA.
DR InterPro: IPR000328; Env_GP41.
DR Pfam: PF00517; GP41; 1.
DR Transmembrane. 1
KW NON_TER 122 122
FT NON_TER 122 122
SQ SEQUENCE 122 AA; 14699 MW; 4B30A0F65B37A3C5 CRC64;

Query Match 57.5%; Score 42; DB 15; Length 122;
Best Local Similarity 52.5%; Pred. No. 94;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNIWW 8
Db 105 EWTSLWN 112
||||:|

RESULT 46
Q8S2U2 Q8S2U2 PRELIMINARY; PRT; 130 AA.
AC Q8S2U2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created);
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update);
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update);
DE Nucleic acid specific glycine-rich protein 1B.
OS Medicago truncatula (barrel medic).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eucosids 1; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
OC NCBI_TaxID=3880;
RN [1]
RP SEQUENCE FROM N.A.
RA Kevei Z., Vinardelli J.M., Kiss G.B., Kondorosi A., Kondorosi E.;
RT "Glycine-rich proteins encoded by a nodule-specific gene family are
RT implicated in different stages of symbiotic nodule development in
RT Medicago.";
RL Mol. Plant Microbe Interact. 0:0-0 (2002).
DR EMBL: AF498991; AAM18950.1; -.
SQ SEQUENCE 130 AA; 13657 MW; A6D4F5F15626A32A CRC64;

Query Match 57.5%; Score 42; DB 10; Length 130;
Best Local Similarity 44.4%; Pred. No. 99;
Matches 4; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 WTNINWWAK 10
Db 86 WGTSMWGR 94
||||:

RESULT 47
Q8UR19 Q8UR19 PRELIMINARY; PRT; 133 AA.
AC Q8UR19;
DT 01-MAR-2002 (TrEMBLrel. 20, Created);
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update);
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update);
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroviridae; Retroviridae; Lentivirus.
OC NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RA Trask S.A., Derdeyn C.A., Fiedli U., Chen Y., Meleth S., Kasolo F.,
RX MEDLINE=21602569; PubMed=11739704;
RX Musonda R., Hunter E., Gao F., Allen S., Hahn B.H.;
RT "Molecular epidemiology of human immunodeficiency virus type 1
RT transmission in a heterosexual cohort of discordant couples in
RT Zambia.";
RL J. Virol. 76:397-405 (2002).
DR EMBL: AF405129; AA56650.1; -.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR GO: GO:0019031; C: viral envelope; IEA.
DR GO: GO:0005198; F: structural molecule activity; IEA.
DR InterPro: IPR000328; Env_GP41.
DR Pfam: PF00517; GP41; 1.
DR Transmembrane. 1
KW NON_TER 1 1
FT NON_TER 133 133
SQ SEQUENCE 133 AA; 16036 MW; 00BAE9738ED5C389 CRC64;

Query Match 57.5%; Score 42; DB 15; Length 133;
Best Local Similarity 71.4%; Pred. No. 1e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 WTNINWW 8
Db 114 WTNLWNW 120
||||:|

RESULT 49
Q41871 Q41871 PRELIMINARY; PRT; 142 AA.
AC Q41871;
DT 01-NOV-1996 (TrEMBLrel. 01, Created);
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update);
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update);
DE Hypothetical protein.
OS Zea mays (Maize).
```

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC PACCAD clade; Panicoideae; Andropogoneae; Zea.
 OX NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Seedlings aerial part;
 RX MEDLINE=89345638; PubMed=2474829;
 RA Jin Y.K., Bennetzen J.L.;
 RT "Structure and coding properties of Bsl, a maize retrovirus-like
 transposon.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:6235-6239 (1989).
 DR EMBL; M25397; AAAG6270.1; -;
 DR PIR; T01976; T03976.
 KW Hypothetical protein.
 SQ SEQUENCE 142 AA; 15828 MW; 492P54531F47E934 CRC64;
 Query Match 57.5%; Score 42; DB 10; Length 142;
 Best Local Similarity 71.4%; Pred. No. 1.1e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 2 WTNIIWW 8
 Db 68 FTNPWW 74
 RESULT 50
 Q32670 PRELIMINARY; PRT; 143 AA.
 ID Q926J0
 AC Q926J0;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE HTH transcriptional regulator [Putative hth transcription
 regulator].
 DE YFGA OR CPN1069 OR CP0781 OR CPB114.
 GN Chlamydia pneumoniae (Chlamydia pneumoniae).
 OS Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 OC NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CWL029;
 RX MEDLINE=99206606; PubMed=10192388;
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 OLinger L., Grimwood J., Davis R.W., Stephens R.S.;
 ET "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 RL Nat. Genet. 21:385-389 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
 RA Lincker K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eiset J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
 pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406 (2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 from Japan and CWL029 from USA.";
 RL Nucleic Acids Res. 28:2311-2314 (2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TW-183;
 RA Geng M.M., Schuhmacher A., Muehldorfer I., Bensch K.W., Schaefer K.P.,
 RA Schneider S., Pohl T., Essig A., Marre R., Melchers K.;

*The genome sequence of Chlamydia pneumoniae TW183 and comparison with
 other Chlamydia strains based on whole genome sequence analysis.*;
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AE001687; AAD19206.1; -;
 DR EMBL; AF002238; AAF38580.1; -;
 DR EMBL; AF002548; BAA99276.1; -;
 DR EMBL; AE017160; AAP99042.1; -;
 DR PIR; B86624; B86624.
 DR PIR; G72000; G72000.
 DR TIGR; CP0781; -;
 KW Complete proteome.
 SQ SEQUENCE 143 AA; 16337 MW; 0BA53P29912ED3D8 CRC64;

Query Match 57.5%; Score 42; DB 16; Length 143;
 Best Local Similarity 66.7%; Pred. No. 1.1e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 WTNIIWW 7
 Db 117 WSNLWW 122

RESULT 51

Q8SCR7 PRELIMINARY; PRT; 144 AA.

AC Q8SCR7;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE PHIKZ245.
 OS Pseudomonas phage phiKZ.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae.
 OC NCBI_TaxID=169683;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21914557; PubMed=11916376;
 RA Mesyanzhinov V.V., Robben J., Grymonprez B., Kostynchenko V.A.,
 RA Bourkalsseva M.V., Sykilinda N.N., Krylov V.N., Volkart G.;
 RT "The genome of bacteriophage phiKZ of Pseudomonas aeruginosa.";
 RL J. Mol. Biol. 317:11-19 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Mesyanzhinov V.V., Robben J., Grymonprez B., Kostynchenko V.A.,
 RA Bourkalsseva M.V., Sykilinda N.N., Krylov V.V., Volkart G.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF399011; AAL83146.1; -;
 SQ SEQUENCE 144 AA; 17414 MW; 201D2613A88738C2 CRC64;

Query Match 57.5%; Score 42; DB 9; Length 144;
 Best Local Similarity 62.5%; Pred. No. 1.1e+02;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNIWW 8
 Db 37 EBTGLWW 44

RESULT 52

O07171 PRELIMINARY; PRT; 144 AA.

AC O07171;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein RV0121C.
 GN RV0121C OR MT0129 OR MTCT418B.03C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OC NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37Rv;

RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekoaia F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Z96071; CAB03449.1; -.
 DR EMBL; AE006925; AAK44353.1; -.
 DR TIGR; B70983; B70983.
 DR TIGR; MT0129; -.
 DR TubercuList; RV0121c; -.
 DR InterPro; IPR009002; FMN binding.
 KW Hypothetical protein; Complete proteome.
 FT CONFLICT 120 120 A -> T (IN REF. 2).
 FT CONFLICT 142 142 W -> R (IN REF. 2).
 SQ SEQUENCE 144 AA; 16045 MW; 96EDB60CE8F38649 CRC64;

 Query Match 57.5%; Score 42; DB 16; Length 144;
 Best Local Similarity 57.1%; Pred. No. 1.1e+02;
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

 QY 1 EWTNIW 7
 Db 86 DWTQLW 92

 RESULT 53
 Q7U2T3 PRELIMINARY; PRT; 144 AA.
 ID Q7U2T3
 AC Q7U2T3
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Conserved hypothetical protein.
 GN M30126C
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AP2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeier K., Canus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkitt R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 DR EMBL; BX248334; CAD92987.1; -.
 KW Complete proteome.
 SQ SEQUENCE 144 AA; 16045 MW; 96EDB60CE8F38649 CRC64;

Query Match 57.5%; Score 42; DB 16; Length 144;
 Best Local Similarity 57.1%; Pred. No. 1.1e+02;
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNIW 7
 Db 86 DWTQLW 92

 RESULT 54
 Q8N8Z4 PRELIMINARY; PRT; 151 AA.
 ID Q8N8Z4
 AC Q8N8Z4
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Hypothetical protein FLJ38668.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
 RA Fukushima Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
 RA Otsuki T., Sato K., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
 RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
 RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K.,
 RA Isogai T.;
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK095987; BAC04666.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 151 AA; 16903 MW; D6BE6902FBI63327 CRC64;

 Query Match 57.5%; Score 42; DB 4; Length 151;
 Best Local Similarity 57.1%; Pred. No. 1.1e+02;
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

 QY 4 NINWAK 10
 Db 106 SVMWKG 112

 RESULT 55
 Q94CI8 PRELIMINARY; PRT; 284 AA.
 ID Q94CI8
 AC Q94CI8
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Glycine-rich protein LeGR1.
 OS Lycopersicon esculentum (tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Lamiales; Solanales; Solanaceae; Solanum.
 OX NCBI_TaxID=4081;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Uthappa M., Muchhal U.S., Baldwin J.C., Raghothama K.G.;
 RT "LeGR1: A new member of glycine-rich proteins from tomato
 RT (Lycopersicon esculentum).";
 RL Physiol. Plantarum 0:0-0(2001).
 DR EMBL; AY026037; AAX08984.1; -.
 DR InterPro; IPR002952; Eggshell.
 DR PRINTS; PR01228; EGGSHLL.
 SQ SEQUENCE 284 AA; 23434 MW; B81A84C247CB9ED8 CRC64;

Query Match 57.5%; Score 42; DB 10; Length 284;
 Best Local Similarity 44.4%; Pred. No. 2.1e+02;
 Matches 4; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 WTNIWMAK 10

```

DR GO: GO:0008460; F:dTDP-glucose 4,6-dehydratase activity; IEA.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR GO: GO:0009225; P:nucleotide-sugar metabolism; IEA.
DR InterPro: IPR002198; dTDP short.
DR InterPro: IPR005888; dTDP gluc dehyd.
DR InterPro: IPR001509; Epimerase_Dh.
DR Pfam: PF01370; Epimerase; 1.
DR TIGRfams: TIGR01181; dTDP gluc dehyd; 1.
DR PROSITE: PS00061; ADH_SHORT; 1.
SQ SEQUENCE 351 AA; 39006 MW; 8AEAO319D8F4BA38 CRC64;

Query Match 57.5%; Score 42; DB 2; Length 351;
Best Local Similarity 60.0%; Pred No: 2.5e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

Qy 1 EW-TNIW 8
Db 323 EWYLDNAWW 332

RESULT 58
Q80U738 PRELIMINARY; PRT; 352 AA.
AC Q80U738;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE DTDP-D-glucose-4,6-dehydratase.
GN RFBB OR ATU4617 OR AGR L 530.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayin T., Levy R., Li M.-J., McLelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tilgney S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RA "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RT Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Iartchouk O., Epp A., Liu F.,
RA Roumieu K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Karkelz B.,
RA Pianagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
DR EMBL; AF009389; AAL4541.1; -
DR EMBL; AF008226; AAK88830.1; -
DR FIR; AE3124; AE3124.
DR FIR; D98163; D98163.
DR GO: GO:0008460; F:dTDP-glucose 4,6-dehydratase activity; IEA.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR GO: GO:0009225; P:nucleotide-sugar metabolism; IEA.
DR InterPro: IPR002198; ADH short.
DR InterPro: IPR005888; dTDP gluc dehyd.
DR InterPro: IPR001509; Epimerase_Dh.

```


DR Pfam: PF01370; Epimerase; 1.
 DR TIGRams: TIGR01181; 4TDP_glu_c dehyd; 1.
 DR PROSITE: PS00061; ADH_SHORT; 1.
 KW Complete proteome.
 SQ SEQUENCE 352 AA; 39089 MW; A80CA11341880D7F CRC64;

Query Match 57.5%; Score 42; DB 16; Length 352;
 Best Local Similarity 60.0%; Pred. No. 2.5e+02;
 Matches 6; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

QY 1 EW--TNIWWW 8
 |||
 324 EWLDAWAWW 333

DR Q33W25 PRELIMINARY; PRT; 374 AA.
 AC Q83W25;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Aca13 protein.
 GN Aca13.
 OS Streptomyces capreolus.
 CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 CC Pseudonocardiales; Actinosynnemataceae; Saccharothrix.
 CX NCBI_TaxID=66854;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL3817;
 RX MEDLINE=95255257; PubMed=7737149;
 RA Barrasa M.I., Tercero J.A., Lacalle R.A., Jimenez A.;
 RT "The arid gene from Streptomyces capreolus encodes a polypeptide of
 RT the ABC-transporters superfamily which confers resistance to the
 RT aminonucleoside antibiotic A201A.";
 RL Eur. J. Biochem. 228:562-569(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL3817;
 RX MEDLINE=97274639; PubMed=9128724;
 RA Barrasa M.I., Tercero J., Jimenez A.;
 RT "The aminonucleoside antibiotic A201A is inactivated by a
 RT phosphotransferase activity from Streptomyces capreolus NRRL 3817, the
 RT producing organism. Isolation and molecular characterization of the
 RT relevant encoding gene and its DNA flanking regions.";
 RL Eur. J. Biochem. 245:54-63(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL3817;
 RX MEDLINE=22310933; PubMed=12423351;
 RA Saugar I., Sanz E., Rubio M., Espinosa J., Jimenez A.;
 RT "Identification of a set of genes involved in the biosynthesis of the
 RT aminonucleoside moiety of antibiotic A201A from Streptomyces
 RT capreolus.";
 RL Eur. J. Biochem. 269:5527-5535(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL3817;
 RA Sanz E., Saugar I., Jimenez A.;
 RT "Cloning and heterologous expression of the antibiotic A201A
 RT biosynthetic gene cluster.";
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X84374; CAD62186.1;
 DR GO: GO:0009058; Pbiocynthesis; IEA.
 DR InterPro: IPR001296; Glyco_transf.1.
 DR Pfam: PF00534; Glycos_transf.1; I.
 SQ SEQUENCE 374 AA; 41652 MW; F653EF132AD9DB56 CRC64;

Query Match 57.5%; Score 42; DB 2; Length 374;
 Best Local Similarity 80.0%; Pred. No. 2.7e+02;
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 NIWWM 8
 |||
 98 NVWWM 102

DR Q82F00 PRELIMINARY; PRT; 439 AA.
 AC Q82F00;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative glycosyl transferase.
 GN SAV4463.
 OS Streptomyces avermitilis.
 CC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 CC Streptomycineae; Streptomycetaceae; Streptomyces.
 CX NCBI_TaxID=33903;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=21477403; PubMed=11572948;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
 RT "Genome sequence of an industrial microorganism Streptomyces
 RT avermitilis: deducing the ability of producing secondary
 RT metabolites.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=22698306; PubMed=12692562;
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
 RA Sakaki Y., Hattori M., Omura S.;
 RT "Complete genome sequence and comparative analysis of the industrial
 RT microorganism Streptomyces avermitilis.";
 RL Nat. Biotechnol. 21:526-531(2003).
 DR EMBL: AP005039; BAC72175.1;
 DR GO: GO:0016740; F:transferase activity; IEA.
 DR InterPro: IPR001173; Glyco_transf.2.
 DR Pfam: PF00535; Glycos_transf.2; I.
 KW Transferase; Complete proteome.
 SQ SEQUENCE 439 AA; 49718 MW; 94B4C638476D8A8F CRC64;

Query Match 57.5%; Score 42; DB 16; Length 439;
 Best Local Similarity 57.1%; Pred. No. 3.1e+02;
 Matches 4; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 WTIWWM 8
 |||
 290 WARSWM 296

DR Q8HQM4 PRELIMINARY; PRT; 460 AA.
 AC Q8HQM4;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE NADH dehydrogenase subunits 4.
 GN ND4.
 OS Lepisosteus oculatus (Spotted gar).
 CC Mitochondrion.
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Semionotiformes; Lepisosteidae;
 CC Lepisosteus.
 CX NCBI_TaxID=7918;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Inoue J., Miya M., Tsukamoto K., Nishida M.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

```

RN SEQUENCE FROM N.A.
RP Inoue J.G., Miya M., Tsukamoto K., Nishida M.;
RT "Basal actinopterygian relationships: a mitogenomic perspective on the
RL phylogeny of the 'ancient fish.'";
DR EMBL; AB042861; BAB40738.1; -.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . .; IEA.
DR InterPro; IPR003918; NADH_oxred4.
DR InterPro; IPR001750; Oxidored_g1.
DR InterPro; IPR000260; Oxidored_g5_N.
DR Pfam; PF00361; oxidored_g1; 1.
DR Pfam; PF01059; oxidored_g5_N; 1.
DR PRINTS; PR01437; NUOXDRDTASE4.
KW Mitochondrion.
SQ SEQUENCE 460 AA; 51533 MW; 59D38EE299461984 CRC64;

Query Match 57.5%; Score 42; DB 8; Length 460;
Best Local Similarity 75.0%; Pred. No. 3.3e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 WTNKIWWA 9
Db 191 WTNKIWWA 198

RESULT 62
QHMM5 PRELIMINARY; PRT; 460 AA.
AC Q8HNM5;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE NADH dehydrogenase subunit 4.
GN ND4.
OS Lepisosteus spatula (Alligator gar) (Atractosteus spatula).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Semionotiformes; Lepisosteidae;
OC Lepisosteus.
OX NCBI_TaxID=7917;
RN [1]
RP SEQUENCE FROM N.A.
RA Inoue J.G.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RA Inoue J.G., Miya M., Tsukamoto K., Nishida M.;
RT "Basal actinopterygian relationships: a mitogenomic perspective on the
RL phylogeny of the 'ancient fish.'";
DR EMBL; AF004355; BAC23993.1; -.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . .; IEA.
DR InterPro; IPR003918; NADH_oxred4.
DR InterPro; IPR001750; Oxidored_g1.
DR InterPro; IPR000260; Oxidored_g5_N.
DR Pfam; PF00361; oxidored_g1; 1.
DR Pfam; PF01059; oxidored_g5_N; 1.
DR PRINTS; PR01437; NUOXDRDTASE4.
KW Mitochondrion.
SQ SEQUENCE 460 AA; 51459 MW; 7C7B2AED4A6488C4 CRC64;

Query Match 57.5%; Score 42; DB 8; Length 460;
Best Local Similarity 75.0%; Pred. No. 3.3e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 WTNKIWWA 9
Db 191 WTNKIWWA 198

us-10-726-148a-15_copy_428_437.rspt
```

```

RESULT 63
P94793 PRELIMINARY; PRT; 466 AA.
ID P94793;
AC P94793;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Phospho-beta-glucosidase (EC 3.2.1.86) (Fragment).
GN PBGA.
OS Fusobacterium mortiferum.
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=850;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25557;
RX MEDLINE=97197528; PubMed=9045824;
RA Thompson J., Robrish S.A., Bouma C.L., Freedberg D.I., Folk J.E.;
RT "Phospho-beta-glucosidase from Fusobacterium mortiferum: purification,
RL cloning, and inactivation by 6-phosphoglucono-delta-lactone.";
DR EMBL; U81184; AAB49339.1; -.
DR HSSP; P11546; 1PBG.
DR GO; GO:0008706; F:6-phospho-beta-glucosidase activity; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001360; Glyco_hydro_1.
DR Pfam; PF00232; Glyco_hydro_1; 1.
DR PRINTS; PR00131; GLHYDRLASE1.
DR ProDom; PD000650; Glyco_hydro_1; 1.
DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 1.
DR PROSITE; PS00653; GLYCOSYL_HYDROL_F1_2; 1.
KW Glycosidase; Hydrolase.
FT NON TER 1
SQ SEQUENCE 466 AA; 53532 MW; 659818CABE610B65 CRC64;

Query Match 57.5%; Score 42; DB 2; Length 466;
Best Local Similarity 80.0%; Pred. No. 3.3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 NIVWW 8
Db 243 NVWW 247

RESULT 64
QHON5 PRELIMINARY; PRT; 544 AA.
ID Q9HON5;
AC Q9HON5;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Glycine betaine transporter.
GN GPUT OR VNG1077G.
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leitbauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.B., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
```

DR EMBL; AE005040; AAG19478.1; -;
 DR PIR; B84264; B84264;
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005215; C:transporter activity; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR000060; BCGT_transporter.
 DR Pfam; PF02028; BCGT; 1.
 DR ProDom; PD01011; BCGT_transporter; 1.
 DR Complete proteome.
 DR SEQUENCE 544 AA; 58168 MW; AASD7266F29A5E47 CRC64;
 Query Match 57.5%; Score 42; DB 17; Length 544;
 Best Local Similarity 60.0%; Pred. No. 3.8e+02;
 Matches 5; Conservative 0; Mismatches 2; Indels 2; Gaps 1;
 Y 2 WTNIW--WNA 9
 D 325 WTGFNAWNA 334
 RESULT 65
 Q90CW2 PRELIMINARY; PRT; 847 AA.
 ID Q90CW2
 AC Q90CW2;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Envelope glycoprotein.
 GN ENV.
 OS Human immunodeficiency virus 1.
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=97TZ08;
 MEDLINE=21395692; PubMed=11504977;
 RA Hoelscher M., Kim B., Maboko L., Mhalu F., von Sonnenburg F.,
 Birk D.L., McCutchan F.E.,
 RA the UNAIDS Network for HIV Isolation Characterization.;
 RT "High proportion of unrelated HIV-1 intersubtype recombinants in the
 RT Mweya region of southwest Tanzania.";
 RL AIDS 15:1461-1470(2001).
 DR EMBL; AF361878; AAK94279.1; -;
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0019028; C:viral capsid; IEA.
 DR GO; GO:0019031; C:viral envelope; IEA.
 DR GO; GO:0005198; P:structural molecule activity; IEA.
 DR InterPro; IPR000328; Env GP41.
 DR Pfam; PF00516; GP120; 1.
 DR Pfam; PF00517; GP41; 1.
 DR AIDS; CoA: protein; Glycoprotein; Polypeptide; Transmembrane.
 DR SEQUENCE 847 AA; 95766 MW; B3C4208FC76937C CRC64;
 Query Match 57.5%; Score 42; DB 15; Length 847;
 Best Local Similarity 71.4%; Pred. No. 5.8e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Y 2 WTNIW--WNA 8
 D 650 WTNIW--WNA 656

RESULT 66
 Q9DVL2 PRELIMINARY; PRT; 854 AA.
 ID Q9DVL2
 AC Q9DVL2;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Envelope protein (Fragment).
 GN ENV OR GP160.
 OS Human immunodeficiency virus 1.

OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J 97DC.KTB147;
 MEDLINE=20499072; PubMed=11044094;
 RA Vidal N., Peeters M., Mulanga-Kabeya C., Nzilambi N., Robertson D.,
 Ilunga W., Sema H., Tshimanga K., Bongo B., Delaporte E.,
 RT "Unprecedented degree of human immunodeficiency virus Type 1 (HIV-1)
 RT group M genetic diversity in the Democratic Republic of Congo suggests
 RT that the HIV-1 pandemic originated in Central Africa.";
 RL J. Virol. 74:110498-10507(2000).
 DR EMBL; AJ401041; CAC15049.1; -;
 DR GO; GO:0031602; C:integral to membrane; IEA.
 DR GO; GO:0031903; C:viral capsid; IEA.
 DR GO; GO:0005198; P:structural molecule activity; IEA.
 DR InterPro; IPR000328; Env GP41.
 DR InterPro; IPR000777; GP120.
 DR Pfam; PF00516; GP120; 1.
 DR Pfam; PF00517; GP41; 1.
 DR AIDS; CoA: protein; Glycoprotein; Polypeptide; Transmembrane.
 DR SEQUENCE 854 AA; 96040 MW; 5A13A29231B0EDCF CRC64;
 Query Match 57.5%; Score 42; DB 15; Length 854;
 Best Local Similarity 50.0%; Pred. No. 5.9e+02;
 Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 Y 1 EWTNIW--WNA 10
 D 663 KNTSLWSWFK 672
 RESULT 67
 Q8WQ32 PRELIMINARY; PRT; 859 AA.
 ID Q8WQ32
 AC Q8WQ32;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Possible mitochondrial endonuclease.
 GN P265.32.
 OS Leishmania major.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5864;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Friedlin;
 RA Robben J., Glynnprez B., Weltjens I., Aert R., Volckaert G.,
 Ivens A.C., Quail M., Rajandream M.A., Barrell B.G.,
 RL Submitted (DSC-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Friedlin;
 RX MEDLINE=98146435; PubMed=9477341;
 RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
 Smith D.F.,
 RT "A physical map of the Leishmania major Friedlin genome.";
 RL Genome Res. 8:135-145(1998).
 DR EMBL; AL359716; CAD19427.1; -;
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003677; P:DNA binding; IEA.
 DR GO; GO:0004519; P:endonuclease activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR000637; At hook.
 DR InterPro; IPR001604; Endonuclease.
 DR Pfam; PF01223; Endonuclease; 1.
 DR SMART; SM00384; AT hook; 1.
 DR SMART; SM00477; NUC; 1.
 DR PROSITE; PS01070; NUCLEASE_NON_SPEC; 1.
 DR Endonuclease.
 DR SEQUENCE 859 AA; 93556 MW; 3B8CC36A785BC7E6 CRC64;

```

Query Match          57.5%; Score 42; DB 5; Length 859;
Best Local Similarity 44.4%; Pred. No. 5.9e+02;
Matches 4; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 WTNINWVWAX 10
DB 201 WPGSWWRR 209

RESULT 68
Q90CV3 PRELIMINARY; PRT; 862 AA.
AC Q90CV3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Envelope glycoprotein.
ENV.
GS Human immunodeficiency virus 1.
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=97TZ09;
RX Hoelscher M., Kim B., Maboko L., Mhalu F., von Sonnenburg F.,
RA Birx D.J., McCutchan F.E.,
RA the UNAIDS Network for HIV Isolation Characterization.;
RT "High proportion of unrelated HIV-1 intersubtype recombinants in the
RL Mbeia region of southwest Tanzania.";
RL AIDS 15:1461-1470(2001).
DR EMBL; AF361879; AAK94288.1; -.
DR GO; GO:0016221; C:integral to membrane; IEA.
DR GO; GO:0013028; C:viral capsid; IEA.
DR GO; GO:0019031; F:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env GP41.
DR InterPro; IPR003777; GP120.
DR Pfam; PF05517; GP41; 1.
DR Pfam; PF05516; GP120; 1.
KW AIDS; Coat protein; Glycoprotein; Polypeptide; Transmembrane.
SQ SEQUENCE 862 AA; 97429 MW; 8FF137E9583D04BC CRC64;

Query Match          57.5%; Score 42; DB 15; Length 862;
Best Local Similarity 62.5%; Pred. No. 5.9e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNINWVW 8
DB 664 EWANLWVW 671

RESULT 69
Q8XFPQ3 PRELIMINARY; PRT; 863 AA.
AC Q8XFPQ3;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative RHS-related transmembrane protein.
ENV.
GS RSP1585 OR RS02144.
OC Ralstonia solanacearum (Pseudomonas solanacearum).
OG Plasmid megaplasmid.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-GM11009;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,

```

```

RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin M., Schiex T.,
RA Siguer F., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT *Genome sequence of the plant pathogen Ralstonia solanacearum.*;
RL Nature 415:497-502(2002).
DR EMBL; AL646085; CAD18736.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR InterPro; IPR006530; YD.
DR Pfam; PF05593; RHS repeat; 7.
DR TIGRFAMs; TIGR01643; YD repeat_2x; 8.
KW Plasmid; Complete proteome.
SQ SEQUENCE 863 AA; 93069 MW; 9B26D294F467AE8B CRC64;

Query Match          57.5%; Score 42; DB 16; Length 863;
Best Local Similarity 55.8%; Pred. No. 5.9e+02;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 WTNINWVWAX 10
DB 706 WTELWTLK 714

RESULT 70
Q9W2J2 PRELIMINARY; PRT; 957 AA.
AC Q9W2J2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CG18375 protein.
GN CG18375.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J.G., Nelson C.R., Miklos G.L.G.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadien E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., McPherson D.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., Moberg A.,
RA Merkulov G., Milshina N.V., Moberg C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,

```

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Krontziller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houch J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattrei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Neelson K.A., Nunoo J.,
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krontziller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003453; AAF46693.2; -;
 DR HSPF; Q13625; 1YCS.
 DR FlyBase; Fgsm0034606; CG18375.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00023; ank; 2.
 DR Pfam; PF00018; SH3; 1.
 DR ProDom; PD00066; SH3; 1.
 DR SMART; SM00248; ANK; 2.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 2.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS50002; SH3; 1.
 KW ANK repeat; Repeat.
 SQ SEQUENCE 997 AA; 107821 MW; E712D400C2C4F03D CRC64;
 Query Match 57.5%; Score 42; DB 5; Length 997;
 Best Local Similarity 71.4%; Pred. No. 6.8e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 4 NIWVWAX 10
 Db 962 NEWVWAX 968
 ID Q86PC3 PRELIMINARY; PRT; 1020 AA.
 AC Q86PC3
 DT 31-JUN-2003 (TrEMBLrel. 24, Created)
 DT 31-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE RE13301P.
 GN CG18375.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; *Drosophila*.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Y;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
 RA George R., Gonzalez M., Guarin H., Krontziller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunoo J., Pacle J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Celniker S.;
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BT003215; AAO24970.1; -;
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00023; ank; 2.
 DR Pfam; PF00018; SH3; 1.
 DR ProDom; PD00066; SH3; 1.
 DR SMART; SM00248; ANK; 2.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 2.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS50002; SH3; 1.
 SQ SEQUENCE 1020 AA; 110434 MW; 42A3AE30EC71787B CRC64;
 Query Match 57.5%; Score 42; DB 5; Length 1020;
 Best Local Similarity 71.4%; Pred. No. 6.9e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 4 NIWVWAX 10
 Db 985 NEWVWAX 991
 ID Q86BG1 PRELIMINARY; PRT; 1069 AA.
 AC Q86BG1
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE CG18375-PB.
 GN CG18375.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; *Drosophila*.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Y;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Houten D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moehref A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pat S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C.M., Bernat B.P., Carlson J.W., Celisner S.E.,
RA Clamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J.,
RA Harris N.L., Krommiller B., Marshall B., Milburn G.H., Richter J.,
RA Russo S., Searle S.M.J., Smith E., Shu S., Smutniak F.,
RA Whitfield E.J., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J.,
RA Lewis S.E.;
RT "Annotation of *Drosophila melanogaster* genome.";
Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003453; AAC01341.1; -;
DR InterPro; IPR002110; ANK.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00023; ank; 2.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PDC00066; SH3; 1.
DR SMART; SM00248; ANK; 2.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00088; ANK_REPEAT; 2.
DR PROSITE; PS00297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00002; SH3; 1.
SQ SEQUENCE 1069 AA; 115158 MW; BF102BC0C44F80DA CRC64;
Query Match 57.5%; Score 42; DB 5; Length 1069;
Best Local Similarity 71.4%; Pred. No. 7.2e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 4 NIWVWAK 10
DB 1034 NEWVWAK 1040
RESULT 73
QYAS583 PRELIMINARY; PRT; 227 AA.
AC QYAS583;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein CC2574.
GN CC2574.

OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nieman W.C., Feldblyum T.V., Taub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Kravén A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Kallenberg S.L., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Utterback T., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of *Caulobacter crescentus*.";
Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005925; AAK24544.1; -;
DR PIR; D87568; D87568.
DR TIGR; CC2574; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 227 AA; 24191 MW; D699A7E875DD8BE5 CRC64;
Query Match 56.8%; Score 41.5; DB 16; Length 227;
Best Local Similarity 45.5%; Pred. No. 2e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 3; Gaps 1;
QY 2 WTNW---WA 9
DB 208 WSSVWMAWA 218
RESULT 74
QY24W1 PRELIMINARY; PRT; 264 AA.
AC QY24W1;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative ABC transporter integral membrane protein.
GN SC03223 OR SCB8.16C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=2196410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2)."
RL Nature 417:141-147(2002).
DR EMBL; AL939115; CAB38594.1; -;
DR PIR; T36315; T36315.
KW Complete proteome.
SQ SEQUENCE 264 AA; 27180 MW; C72BD78D1D566EF0 CRC64;
Query Match 56.8%; Score 41.5; DB 16; Length 264;
Best Local Similarity 46.2%; Pred. No. 2.3e+02;
Matches 5; Conservative 0; Mismatches 2; Indels 5; Gaps 1;
QY 1 EWTNIW-----W 8
DB 25 EWTNWSVRATW 37

RESULT 75

Q7SVF0 PRELIMINARY; PRT; 117 AA.

AC Q7SVF0; 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Envelope glycoprotein gp41 domain (Fragment).

GN ENV.

OS Human immunodeficiency virus 1.

OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

CX KCB1_TaxID=11676;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=99NG804;

RA Zeh C., Agwale S.M., Robbins K.E., Odama L., Sani-Gwarco N.,

RA Gboun M.S., Pieniazek D., Wambebe C., Folks T.M., Kalish M.L.;

RT "Presence of HIV-2 subtype B in single and HIV-2 subtype A in dual

RT HIV-1/HIV-2 infections in Nigeria."

RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY102834; AAM51929.1; -

KW Envelope protein.

FT NON TER 1

FT NON TER 117

SQ SEQUENCE 117 AA; 14111 MW; 5BF2G6BEC093F54F CRC64;

Query Match 56.2%; Score 41; DB 15; Length 117;

Best Local Similarity 50.0%; Pred.No. 1.2e+02;

Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EWYN¹WWAX 10

Db 105 KWANLWVFK 114

Search completed: June 9, 2004, 13:11:17

Job time : 71 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2004, 18:08:32 ; Search time 23 Seconds
(without alignments)
22.446 Million cell updates/sec

Title: US-10-726-148a-15_COPY_428_437

Perfect score: 73
Sequence: 1 EWTNIWWAK 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 segs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 3%

Maximum Match 100%

Listing first 300 summaries

Database : Issued Patents AA.*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	60.3	435	4	US-09-489-039A-7287
2	43.5	59.6	154	4	US-09-252-991A-29786
3	43	58.9	37	4	US-09-779-451-68
4	43	58.9	47	4	US-09-779-451-66
5	42	57.5	8	3	US-09-082-279B-1509
6	42	57.5	8	4	US-09-315-304B-1663
7	42	57.5	8	4	US-09-834-784-1509
8	42	57.5	8	4	US-09-350-641C-1664
9	42	57.5	143	4	US-09-138-452A-6
10	42	57.5	296	4	US-09-540-236-2118
11	41	56.2	95	4	US-09-489-039A-10296
12	41	56.2	416	3	US-09-333-208-2
13	41	56.2	416	3	US-09-333-254-2
14	41	56.2	416	4	US-09-133-270-2
15	40	54.8	1754	1	US-07-745-206A-13
16	40	54.8	1754	2	US-08-311-363-13
17	39.5	54.1	199	4	US-09-673-395A-570
18	39	53.4	36	4	US-09-779-451-74
19	39	53.4	45	4	US-09-779-451-72
20	39	53.4	60	4	US-08-630-315A-112
21	39	53.4	71	4	US-09-252-991A-20055
22	39	53.4	92	4	US-09-621-976-5285
23	39	53.4	144	4	US-08-630-915A-6
24	39	53.4	183	4	US-09-543-681A-5268
25	39	53.4	194	4	US-08-311-731A-306
26	39	53.4	240	4	US-09-252-991A-19090
27	39	53.4	243	3	US-08-965-056-73
28	39	53.4	243	4	US-09-479-645A-142
29	39	53.4	282	4	US-09-247-890-16
30	39	53.4	282	4	US-09-724-969-16
31	39	53.4	282	4	US-09-724-852-16
32	39	53.4	351	4	US-08-817-441-47
33	39	53.4	715	4	US-09-462-917A-134
34	39	53.4	862	4	US-09-206-551-15
35	39	53.4	877	4	US-08-817-441-102
36	39	53.4	912	3	US-08-817-785-2
37	39	53.4	912	4	US-09-641-318-2
38	39	53.4	912	4	US-09-817-464-2
39	39	53.4	912	5	PCT-US91-09422-19
40	39	53.4	1005	2	US-08-935-450-2
41	39	53.4	1005	4	US-09-338-123-2
42	38.5	52.7	20	3	US-08-825-852-49
43	38.5	52.7	20	3	US-09-052-888-50
44	38.5	52.7	20	4	US-09-723-890-50
45	38.5	52.7	20	4	US-09-723-901-50
46	38.5	52.7	20	4	US-09-723-547-50
47	38.5	52.7	20	4	US-09-724-127-50
48	38.5	52.7	20	4	US-09-723-931-50
49	38.5	52.7	20	4	US-09-724-114-50
50	38.5	52.7	20	4	US-09-723-913-50
51	38.5	52.7	294	4	US-09-424-349A-2
52	38.5	52.7	294	4	US-09-424-349A-9
53	38.5	52.7	294	8	US-09-082-279B-1499
54	38	52.1	8	3	US-09-082-279B-1507
55	38	52.1	8	4	US-09-315-304B-1653
56	38	52.1	8	4	US-09-834-784-1499
57	38	52.1	8	4	US-09-350-641C-1573
58	38	52.1	8	4	US-09-350-641C-1573
59	38	52.1	8	4	US-09-350-641C-1573
60	38	52.1	8	4	US-09-350-641C-1573
61	38	52.1	8	4	US-09-350-641C-1573
62	38	52.1	9	4	US-09-315-304B-1564
63	38	52.1	9	4	US-09-315-304B-1580
64	38	52.1	9	4	US-09-350-325-24
65	38	52.1	9	4	US-09-350-325-35
66	38	52.1	9	4	US-09-350-325-44
67	38	52.1	10	4	US-09-350-641C-1573
68	38	52.1	10	4	US-09-350-641C-1573
69	38	52.1	10	4	US-09-315-304B-1573
70	38	52.1	10	4	US-09-315-304B-1584
71	38	52.1	10	4	US-09-350-325-35
72	38	52.1	10	4	US-09-350-325-44
73	38	52.1	10	4	US-09-350-641C-1573
74	38	52.1	10	4	US-09-350-641C-1573
75	38	52.1	10	4	US-09-350-641C-1573
76	38	52.1	10	4	US-09-350-641C-1573
77	38	52.1	32	3	US-09-082-279B-1193
78	38	52.1	32	4	US-09-315-304B-1193
79	38	52.1	32	4	US-09-834-784-1193
80	38	52.1	32	4	US-09-515-965A-1193
81	38	52.1	32	4	US-09-350-641C-1193
82	38	52.1	36	1	US-08-073-028-4
83	38	52.1	36	3	US-08-486-099-4
84	38	52.1	36	3	US-08-360-107A-4
85	38	52.1	36	3	US-08-484-223B-4
86	38	52.1	36	3	US-08-475-668A-4
87	38	52.1	36	3	US-08-485-551A-4
88	38	52.1	36	3	US-08-471-913A-4
89	38	52.1	36	3	US-08-554-616-4
90	38	52.1	36	3	US-08-485-264A-4
91	38	52.1	36	3	US-09-082-279B-1026
92	38	52.1	36	3	US-09-082-279B-1027
93	38	52.1	36	3	US-09-082-279B-1028
94	38	52.1	36	3	US-09-082-279B-1358
95	38	52.1	36	3	US-09-082-279B-1359
96	38	52.1	36	4	US-08-474-349A-4
97	38	52.1	36	4	US-09-315-304B-1026
98	38	52.1	36	4	US-09-315-304B-1027
99	38	52.1	36	4	US-09-315-304B-1028
100	38	52.1	36	4	US-09-315-304B-1358

Sequence 142, App
Sequence 16, Appl
Sequence 16, Appl
Sequence 16, Appl
Sequence 47, Appl
Sequence 134, Appl
Sequence 15, Appl
Sequence 182, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 19, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 49, Appl
Sequence 50, Appl
Sequence 50, Appl
Sequence 50, Appl
Sequence 50, Appl
Sequence 50, Appl
Sequence 50, Appl
Sequence 2, Appl
Sequence 9, Appl
Sequence 1499, Ap
Sequence 1507, Ap
Sequence 1653, Ap
Sequence 1499, Ap
Sequence 1507, Ap
Sequence 1654, Ap
Sequence 1662, Ap
Sequence 1564, Ap
Sequence 1580, Ap
Sequence 24, Appl
Sequence 40, Appl
Sequence 1564, Ap
Sequence 1580, Ap
Sequence 1573, Ap
Sequence 1573, Ap
Sequence 1584, Ap
Sequence 1193, Ap
Sequence 1193, Ap
Sequence 1193, Ap
Sequence 1193, Ap
Sequence 1193, Ap
Sequence 4, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 1026, Ap
Sequence 1027, Ap
Sequence 1028, Ap
Sequence 1358, Ap
Sequence 1359, Ap
Sequence 4, Appl
Sequence 1026, Ap
Sequence 1027, Ap
Sequence 1028, Ap
Sequence 1358, Ap

101	38	52.1	36	4	US-09-315-304B-1359	Sequence 1359, Ap	174	38	52.1	39	4	US-09-515-965A-1116	Sequence 1116, Ap
102	38	52.1	36	4	US-09-315-304B-1440	Sequence 1440, Ap	175	38	52.1	39	4	US-09-515-965A-1131	Sequence 1131, Ap
103	38	52.1	36	4	US-09-315-304B-1441	Sequence 1441, Ap	176	38	52.1	39	4	US-09-515-965A-1150	Sequence 1150, Ap
104	38	52.1	36	4	US-08-255-308A-4	Sequence 4, Appl	177	38	52.1	39	4	US-09-515-965A-1151	Sequence 1151, Ap
105	38	52.1	36	4	US-09-973-952-4	Sequence 4, Appl	178	38	52.1	39	4	US-09-515-965A-1152	Sequence 1152, Ap
106	38	52.1	36	4	US-08-470-896-4	Sequence 4, Appl	179	38	52.1	39	4	US-09-515-965A-1165	Sequence 1165, Ap
107	38	52.1	36	4	US-08-485-546A-4	Sequence 4, Appl	180	38	52.1	39	4	US-09-515-965A-1270	Sequence 1270, Ap
108	38	52.1	36	4	US-09-834-784-1026	Sequence 1026, Ap	181	38	52.1	39	4	US-09-515-965A-1273	Sequence 1273, Ap
109	38	52.1	36	4	US-09-834-784-1027	Sequence 1027, Ap	182	38	52.1	39	4	US-09-515-965A-1432	Sequence 1432, Ap
110	38	52.1	36	4	US-09-834-784-1028	Sequence 1028, Ap	183	38	52.1	39	4	US-09-350-641C-1098	Sequence 1098, Ap
111	38	52.1	36	4	US-09-834-784-1358	Sequence 1358, Ap	184	38	52.1	39	4	US-09-350-641C-1116	Sequence 1116, Ap
112	38	52.1	36	4	US-09-834-784-1359	Sequence 1359, Ap	185	38	52.1	39	4	US-09-350-641C-1131	Sequence 1131, Ap
113	38	52.1	36	4	US-09-779-451-53	Sequence 53, Appl	186	38	52.1	39	4	US-09-350-641C-1150	Sequence 1150, Ap
114	38	52.1	36	4	US-09-515-965A-1026	Sequence 1026, Ap	187	38	52.1	39	4	US-09-350-641C-1151	Sequence 1151, Ap
115	38	52.1	36	4	US-09-515-965A-1027	Sequence 1027, Ap	188	38	52.1	39	4	US-09-350-641C-1152	Sequence 1152, Ap
116	38	52.1	36	4	US-09-515-965A-1028	Sequence 1028, Ap	189	38	52.1	39	4	US-09-350-641C-1270	Sequence 1270, Ap
117	38	52.1	36	4	US-09-515-965A-1038	Sequence 1038, Ap	190	38	52.1	39	4	US-09-350-641C-1273	Sequence 1273, Ap
118	38	52.1	36	4	US-09-515-965A-1358	Sequence 1358, Ap	191	38	52.1	39	4	US-09-350-641C-1432	Sequence 1432, Ap
119	38	52.1	36	4	US-09-515-965A-1359	Sequence 1359, Ap	192	38	52.1	39	4	US-09-350-641C-1437	Sequence 1437, Ap
120	38	52.1	36	4	US-09-515-965A-1440	Sequence 1440, Ap	193	38	52.1	40	4	US-09-315-304B-1437	Sequence 1437, Ap
121	38	52.1	36	4	US-09-515-965A-1441	Sequence 1441, Ap	194	38	52.1	40	4	US-09-315-304B-1438	Sequence 1438, Ap
122	38	52.1	36	4	US-09-350-641C-1026	Sequence 1026, Ap	195	38	52.1	40	4	US-09-315-304B-1439	Sequence 1439, Ap
123	38	52.1	36	4	US-09-350-641C-1027	Sequence 1027, Ap	196	38	52.1	40	4	US-09-315-304B-1442	Sequence 1442, Ap
124	38	52.1	36	4	US-09-350-641C-1028	Sequence 1028, Ap	197	38	52.1	40	4	US-09-315-304B-1443	Sequence 1443, Ap
125	38	52.1	36	4	US-09-350-641C-1358	Sequence 1358, Ap	198	38	52.1	40	4	US-09-515-965A-1438	Sequence 1438, Ap
126	38	52.1	36	4	US-09-350-641C-1359	Sequence 1359, Ap	199	38	52.1	40	4	US-09-515-965A-1439	Sequence 1439, Ap
127	38	52.1	36	4	US-09-350-641C-1440	Sequence 1440, Ap	200	38	52.1	40	4	US-09-515-965A-1442	Sequence 1442, Ap
128	38	52.1	37	3	US-09-350-641C-1441	Sequence 1441, Ap	201	38	52.1	40	4	US-09-515-965A-1443	Sequence 1443, Ap
129	38	52.1	37	3	US-09-315-304B-1117	Sequence 1117, Ap	202	38	52.1	40	4	US-09-515-965A-1438	Sequence 1438, Ap
130	38	52.1	37	4	US-09-834-784-1117	Sequence 1117, Ap	203	38	52.1	40	4	US-09-515-965A-1439	Sequence 1439, Ap
131	38	52.1	37	4	US-09-515-965A-1117	Sequence 1117, Ap	204	38	52.1	40	4	US-09-350-641C-1438	Sequence 1438, Ap
132	38	52.1	37	4	US-09-350-641C-1117	Sequence 1117, Ap	205	38	52.1	40	4	US-09-350-641C-1439	Sequence 1439, Ap
133	38	52.1	38	3	US-09-350-641C-1117	Sequence 1117, Ap	206	38	52.1	40	4	US-09-350-641C-1442	Sequence 1442, Ap
134	38	52.1	38	3	US-09-350-641C-1117	Sequence 1117, Ap	207	38	52.1	40	4	US-09-350-641C-1443	Sequence 1443, Ap
135	38	52.1	38	3	US-09-350-641C-1117	Sequence 1117, Ap	208	38	52.1	40	4	US-09-350-641C-1438	Sequence 1438, Ap
136	38	52.1	38	4	US-09-315-304B-1166	Sequence 1166, Ap	209	38	52.1	46	3	US-09-082-279B-1089	Sequence 1089, Ap
137	38	52.1	38	4	US-09-315-304B-1167	Sequence 1167, Ap	210	38	52.1	46	3	US-09-082-279B-1089	Sequence 1089, Ap
138	38	52.1	38	4	US-09-834-784-1166	Sequence 1166, Ap	211	38	52.1	46	4	US-09-834-784-1089	Sequence 1089, Ap
139	38	52.1	38	4	US-09-515-965A-1166	Sequence 1166, Ap	212	38	52.1	46	4	US-09-779-451-51	Sequence 51, Appl
140	38	52.1	38	4	US-09-515-965A-1167	Sequence 1167, Ap	213	38	52.1	46	4	US-09-515-965A-1089	Sequence 1089, Ap
141	38	52.1	38	4	US-09-515-965A-1167	Sequence 1167, Ap	214	38	52.1	46	4	US-09-350-641C-1089	Sequence 1089, Ap
142	38	52.1	38	4	US-09-350-641C-1166	Sequence 1166, Ap	215	38	52.1	48	3	US-09-082-279B-1090	Sequence 1090, Ap
143	38	52.1	38	4	US-09-350-641C-1167	Sequence 1167, Ap	216	38	52.1	48	4	US-09-315-304B-1090	Sequence 1090, Ap
144	38	52.1	39	3	US-09-082-279B-1133	Sequence 1098, Ap	217	38	52.1	48	4	US-09-834-784-1090	Sequence 1090, Ap
145	38	52.1	39	3	US-09-082-279B-1133	Sequence 1131, Ap	218	38	52.1	48	4	US-09-515-965A-1090	Sequence 1090, Ap
146	38	52.1	39	3	US-09-082-279B-1133	Sequence 1131, Ap	219	38	52.1	48	4	US-09-350-641C-1090	Sequence 1090, Ap
147	38	52.1	39	3	US-09-082-279B-1133	Sequence 1150, Ap	220	38	52.1	48	4	US-09-350-641C-1090	Sequence 1090, Ap
148	38	52.1	39	3	US-09-082-279B-1133	Sequence 1151, Ap	221	38	52.1	96	4	US-09-489-039A-10258	Sequence 10258, A
149	38	52.1	39	3	US-09-082-279B-1133	Sequence 1152, Ap	222	38	52.1	138	4	US-09-489-039A-10385	Sequence 10385, A
150	38	52.1	39	3	US-09-082-279B-1133	Sequence 1165, Ap	223	38	52.1	138	4	US-09-570-921-18	Sequence 18, Appl
151	38	52.1	39	3	US-09-082-279B-1133	Sequence 1270, Ap	224	38	52.1	138	4	US-09-570-921-23	Sequence 23, Appl
152	38	52.1	39	3	US-09-082-279B-1133	Sequence 1273, Ap	225	38	52.1	138	4	US-09-570-921-23	Sequence 23, Appl
153	38	52.1	39	3	US-09-082-279B-1133	Sequence 1432, Ap	226	38	52.1	143	4	US-09-252-991A-20117	Sequence 20117, A
154	38	52.1	39	4	US-09-315-304B-1038	Sequence 1098, Ap	227	38	52.1	188	4	US-09-543-681A-7763	Sequence 7763, A
155	38	52.1	39	4	US-09-315-304B-1116	Sequence 1116, Ap	228	38	52.1	194	4	US-09-489-039A-14065	Sequence 14065, A
156	38	52.1	39	4	US-09-315-304B-1131	Sequence 1131, Ap	229	38	52.1	237	3	US-09-388-353-643	Sequence 643, App
157	38	52.1	39	4	US-09-315-304B-1150	Sequence 1150, Ap	230	38	52.1	237	3	US-08-488-551B-643	Sequence 643, App
158	38	52.1	39	4	US-09-315-304B-1151	Sequence 1151, Ap	231	38	52.1	244	4	US-09-328-352-6861	Sequence 6861, Ap
159	38	52.1	39	4	US-09-315-304B-1152	Sequence 1152, Ap	232	38	52.1	268	3	US-09-328-352-6861	Sequence 6861, Ap
160	38	52.1	39	4	US-09-315-304B-1155	Sequence 1155, Ap	233	38	52.1	268	3	US-08-965-056-68	Sequence 68, Appl
161	38	52.1	39	4	US-09-315-304B-1270	Sequence 1270, Ap	234	38	52.1	268	3	US-08-965-056-70	Sequence 70, Appl
162	38	52.1	39	4	US-09-315-304B-1273	Sequence 1273, Ap	235	38	52.1	268	3	US-08-965-056-72	Sequence 72, Appl
163	38	52.1	39	4	US-09-315-304B-1432	Sequence 1432, Ap	236	38	52.1	269	3	US-08-965-056-25	Sequence 25, Appl
164	38	52.1	39	4	US-09-834-784-1098	Sequence 1098, Ap	237	38	52.1	269	3	US-08-965-056-33	Sequence 33, Appl
165	38	52.1	39	4	US-09-834-784-1116	Sequence 1116, Ap	238	38	52.1	270	3	US-08-965-056-67	Sequence 67, Appl
166	38	52.1	39	4	US-09-834-784-1131	Sequence 1131, Ap	239	38	52.1	300	4	US-09-489-039A-14307	Sequence 14307, A
167	38	52.1	39	4	US-09-834-784-1150	Sequence 1150, Ap	240	38	52.1	314	1	US-08-118-270-26	Sequence 26, Appl
168	38	52.1	39	4	US-09-834-784-1151	Sequence 1151, Ap	241	38	52.1	314	5	PCT-US93-08528-26	Sequence 26, Appl
169	38	52.1	39	4	US-09-834-784-1152	Sequence 1152, Ap	242	38	52.1	344	4	US-09-489-039A-11458	Sequence 11458, A
170	38	52.1	39	4	US-09-834-784-1165	Sequence 1165, Ap	243	38	52.1	370	4	US-09-252-991A-31298	Sequence 31298, A
171	38	52.1	39	4	US-09-834-784-1270	Sequence 1270, Ap	244	38	52.1	416	4	US-09-252-991A-23400	Sequence 23400, A
172	38	52.1	39	4	US-09-834-784-1273	Sequence 1273, Ap	245	38	52.1	433	3	US-08-691-563C-87	Sequence 87, Appl
173	38	52.1	39	4	US-09-834-784-1432	Sequence 1432, Ap	246	38	52.1	433	4	US-09-374-766-87	Sequence 87, Appl
					Sequence 1432, Ap							US-08-979-847B-81	Sequence 81, Appl
					Sequence 1432, Ap							US-09-543-681A-6490	Sequence 6490, Ap

247 38 52.1 685 4 US-09-489-039A-12981
248 38 52.1 865 3 US-07-956-483-13
249 38 52.1 887 3 US-08-472-240A-4
250 37 50.7 8 3 US-09-082-279B-1475
251 37 50.7 8 3 US-09-082-279B-1490
252 37 50.7 8 3 US-09-082-279B-1494
253 37 50.7 8 3 US-09-082-279B-1502
254 37 50.7 8 4 US-09-315-304B-1629
255 37 50.7 8 4 US-09-315-304B-1644
256 37 50.7 8 4 US-09-315-304B-1648
257 37 50.7 8 4 US-09-315-304B-1656
258 37 50.7 8 4 US-09-315-304B-1661
259 37 50.7 8 4 US-09-834-784-1475
260 37 50.7 8 4 US-09-834-784-1490
261 37 50.7 8 4 US-09-834-784-1494
262 37 50.7 8 4 US-09-834-784-1502
263 37 50.7 8 4 US-09-350-641C-1630
264 37 50.7 8 4 US-09-350-641C-1645
265 37 50.7 8 4 US-09-350-641C-1649
266 37 50.7 8 4 US-09-350-641C-1657
267 37 50.7 9 4 US-09-315-304B-1582
268 37 50.7 9 4 US-09-350-325-42
269 37 50.7 9 4 US-09-350-641C-1582
270 37 50.7 10 4 US-09-315-304B-1578
271 37 50.7 10 4 US-09-350-325-38
272 37 50.7 10 4 US-09-350-641C-1578
273 37 50.7 11 1 US-08-465-391A-328
274 37 50.7 11 2 US-08-464-538B-326
275 37 50.7 11 2 US-08-463-078E-42
276 37 50.7 11 3 US-08-915-314-61
277 37 50.7 11 4 US-09-030-619-56
278 37 50.7 11 4 US-09-667-486-61
279 37 50.7 11 4 US-09-428-082B-232
280 37 50.7 11 4 US-09-428-082B-251
281 37 50.7 13 3 US-08-486-098-172
282 37 50.7 13 3 US-08-484-223B-172
283 37 50.7 13 3 US-08-919-597-172
284 37 50.7 13 3 US-08-475-668A-172
285 37 50.7 13 3 US-08-485-551A-172
286 37 50.7 13 3 US-08-471-913A-172
287 37 50.7 13 3 US-08-485-264A-172
288 37 50.7 13 3 US-09-082-279B-530
289 37 50.7 13 4 US-08-474-349A-172
290 37 50.7 13 4 US-09-315-304B-530
291 37 50.7 13 4 US-08-973-952-25
292 37 50.7 13 4 US-08-470-896-172
293 37 50.7 13 4 US-08-485-546A-172
294 37 50.7 13 4 US-09-834-784-530
295 37 50.7 13 4 US-09-515-965A-530
296 37 50.7 13 4 US-09-350-641C-530
297 37 50.7 38 3 US-09-082-279B-1169
298 37 50.7 38 3 US-09-082-279B-1170
299 37 50.7 38 4 US-09-315-304B-1169
300 37 50.7 38 4 US-09-315-304B-1170

ALIGNMENTS

RESULT 1
US-09-489-039A-7287
; Sequence 7287, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 7287
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7287

Query Match 60.3%; Score 44; DB 4; Length 435;
Best Local Similarity 50.0%; Pred. No. 50;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 WTNINWNA 9
| : ||| :
Db 333 WPHTWNWS 340

RESULT 2

US-09-252-991A-29786
; Sequence 29786, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 29786

; LENGTH: 154

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-29786

Query Match 59.6%; Score 43.5; DB 4; Length 154;
Best Local Similarity 58.3%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

QY 1 EWTNI---WWWA 9
||| |||||
Db 101 EWAAISAAWWWA 112

RESULT 3

US-09-779-451-68
; Sequence 68, Application US/09779451
; Patent No. 6605427
; GENERAL INFORMATION:

; APPLICANT: Wild, Carl T.

; APPLICANT: Allaway, Graham P.

; TITLE OF INVENTION: Assay for Detection of Viral Fusion Inhibitors

; FILE REFERENCE: 1900.030003

; CURRENT APPLICATION NUMBER: US/09/779,451

; CURRENT FILING DATE: 2001-08-17

; PRIOR APPLICATION NUMBER: US 60/235,901

; PRIOR FILING DATE: 2000-09-28

; PRIOR APPLICATION NUMBER: US 60/181,543

; PRIOR FILING DATE: 2000-02-10

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 68

; LENGTH: 37

; TYPE: PRT

; ORGANISM: Human immunodeficiency virus type 1

US-09-779-451-68

Query Match 58.9%; Score 43; DB 4; Length 37;
Best Local Similarity 62.5%; Pred. No. 5.5;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 3;

```
Qy 1 EWTNIWWW 8
Db 28 KWTNLWNW 35

RESULT 4
US-09-779-451-66
; Sequence 66, Application US/09779451
; Patent No. 6605427
; GENERAL INFORMATION:
; APPLICANT: Wild, Carl T.
; TITLE OF INVENTION: Assay for Detection of Viral Fusion Inhibitors
; FILE REFERENCE: 1900.0300CC3
; CURRENT APPLICATION NUMBER: US/09/779,451
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 60/235,901
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/181,543
; PRIOR FILING DATE: 2000-02-10
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 66
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-779-451-66

Query Match 58.9%; Score 43; DB 4; Length 47;
Best Local Similarity 62.5%; Pred. No. 7;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EWTNIWWW 8
Db 38 KWTNLWNW 45

RESULT 5
US-09-082-279B-1509
; Sequence 1509, Application US/09082279B
; Patent No. 6258782
; GENERAL INFORMATION:
; APPLICANT: Barney, Shawn
; APPLICANT: Guthrie, Kelly
; APPLICANT: Merutka, Gene
; APPLICANT: Anwer, Mohamed
; APPLICANT: Lambert, Dennis
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED
; FILE REFERENCE: 7872-043
; CURRENT APPLICATION NUMBER: US/09/082,279B
; CURRENT FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1515
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1509
; LENGTH: 8
; TYPE: PRT
; ORGANISM: HIV-1
US-09-082-279B-1509

Query Match 57.5%; Score 42; DB 3; Length 8;
Best Local Similarity 71.4%; Pred. No. 3e+05;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 WTNIIWWW 8
Db 1 WTNLWNW 7

RESULT 6
US-09-315-304B-1663
; Sequence 1663, Application US/09315304B
; Patent No. 6348568
; GENERAL INFORMATION:
; APPLICANT: Barney, S.
; APPLICANT: Guthrie, K.
; APPLICANT: Merutka, G.
; APPLICANT: Anwer, M.

Qy 2 WTNIIWWW 8
Db 1 WTNLWNW 7

RESULT 7
US-09-834-784-1509
; Sequence 1509, Application US/09834784
; Patent No. 6562787
; GENERAL INFORMATION:
; APPLICANT: Barney, Shawn
; APPLICANT: Guthrie, Kelly
; APPLICANT: Merutka, Gene
; APPLICANT: Anwer, Mohamed
; APPLICANT: Lambert, Dennis
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED
; FILE REFERENCE: 7872-043
; CURRENT APPLICATION NUMBER: US/09/834,784
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 09/082,279
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1515
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1529
; LENGTH: 8
; TYPE: PRT
; ORGANISM: HIV-1
US-09-834-784-1509

Query Match 57.5%; Score 42; DB 4; Length 8;
Best Local Similarity 71.4%; Pred. No. 3e+05;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 WTNIIWWW 8
Db 1 WTNLWNW 7

RESULT 8
US-09-350-641C-1664
; Sequence 1664, Application US/09350641C
; Patent No. 6656906
; GENERAL INFORMATION:
; APPLICANT: Barney, S.
; APPLICANT: Guthrie, K.
; APPLICANT: Merutka, G.
; APPLICANT: Anwer, M.

Qy 2 WTNIIWWW 8
Db 1 WTNLWNW 7

RESULT 9
US-09-315-304B-1663
; Sequence 1663, Application US/09315304B
; Patent No. 6348568
; GENERAL INFORMATION:
; APPLICANT: Barney, S.
; APPLICANT: Guthrie, K.
; APPLICANT: Merutka, G.
; APPLICANT: Anwer, M.
```

APPLICANT: Lambert, D.
TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED PHARMACOKINETIC
PROPERTIES
FILE REFERENCE: 7872-067
CURRENT APPLICATION NUMBER: US/09/350,641C
CURRENT FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: 09/315,304
PRIOR FILING DATE: 1999-05-20
PRIOR APPLICATION NUMBER: 09/082,279
PRIOR FILING DATE: 1998-05-20
NUMBER OF SEQ ID NOS: 1757
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1664
LENGTH: 8
TYPE: PRT
ORGANISM: HIV-1
US-09-350-641C-1664

Query Match 57.5%; Score 42; DB 4; Length 9;
Best Local Similarity 71.4%; Pred. No. 36+05;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 WTNIWWW 8
DB 1 WTNLAW 7

RESULT 9
US-09-198-452A-6

Sequence 6, Application US/09198452A
Patent No. 6559294
GENERAL INFORMATION:
APPLICANT: Griffiths, R.
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 6
LENGTH: 143
TYPE: PRT
ORGANISM: Chlamydia pneumoniae
US-09-198-452A-6

Query Match 57.5%; Score 42; DB 4; Length 143;
Best Local Similarity 66.7%; Pred. No. 30;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 WTNIWW 7
DB 117 WSNLW 122

RESULT 10
US-09-540-236-2118

Sequence 2118, Application US/09540236
Patent No. 6673910
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
FILE REFERENCE: 2709.2005-001
CURRENT APPLICATION NUMBER: US/09/540,236
CURRENT FILING DATE: 2000-04-04
NUMBER OF SEQ ID NOS: 3840
SEQ ID NO 2118
LENGTH: 296
TYPE: PRT
ORGANISM: M. catarrhalis
US-09-540-236-2118

Query Match 57.5%; Score 42; DB 4; Length 296;
Best Local Similarity 62.5%; Pred. No. 64;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WTNIWWW 9
DB 86 WTWVFWW 93

RESULT 11
US-09-489-039A-10296
Sequence 10296, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 10296
LENGTH: 95
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10296

Query Match 56.2%; Score 41; DB 4; Length 95;
Best Local Similarity 71.4%; Pred. No. 28;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 NIWWWAK 10
DB 5 NEWWWGK 11

RESULT 12
US-09-533-208-2

Sequence 2, Application US/09333208A
Patent No. 6225089
GENERAL INFORMATION:
APPLICANT: Chen, Kevin K.
TITLE OF INVENTION: A Putative Solvent/Antibiotic Resistant Gene from Pseudomonas Mendocina
FILE REFERENCE: CL-1160-C
CURRENT APPLICATION NUMBER: US/09/333,208A
CURRENT FILING DATE: 1999-06-15
EARLIER APPLICATION NUMBER: 09/183,270
PRIOR FILING DATE: 1998-10-30
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Microsoft Office 97
SEQ ID NO 2
LENGTH: 416
TYPE: PRT
ORGANISM: Pseudomonas mendocina KR-1
US-09-333-208-2

Query Match 56.2%; Score 41; DB 3; Length 416;
Best Local Similarity 57.1%; Pred. No. 1.3e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNIWW 7
DB 220 EWRVFWW 226

RESULT 13
US-09-333-254-2

Sequence 2, Application US/09333254A
Patent No. 6235882
GENERAL INFORMATION:

APPLICANT: Chen, Kevin K.
TITLE OF INVENTION: A Gene Encoding a Putative Efflux Protein for Solvents/
FILE REFERENCE: CL-1160-D
CURRENT APPLICATION NUMBER: US/09/333,254A
CURRENT FILING DATE: 1999-06-15
EARLIER APPLICATION NUMBER: 08/961,738
EARLIER FILING DATE: 1997-10-31
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Microsoft Office 97
SEQ ID NO 2
LENGTH: 416
TYPE: PRT
ORGANISM: Pseudomonas mendocina KR-1
US-09-333-254-2

Query Match 56.2%; Score 41; DB 3; Length 416;
Best Local Similarity 57.1%; Pred. No. 1.3e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 EWTNIWW 7
DB 220 EWSVW 226

RESULT 14
US-09-183-270-2
Sequence 2, Application US/09183270B
Patent No. 6410265
GENERAL INFORMATION:
APPLICANT: Chen, Kevin K.
TITLE OF INVENTION: A Putative Solvent/Antibiotic Resistant Gene
FILE REFERENCE: CL-1160-A
CURRENT APPLICATION NUMBER: US/09/183,270B
CURRENT FILING DATE: 1998-10-30
EARLIER APPLICATION NUMBER: 08/961,738
EARLIER FILING DATE: 1997-10-31
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Microsoft Office 97
SEQ ID NO 2
LENGTH: 416
TYPE: PRT
ORGANISM: Pseudomonas mendocina KR-1
US-09-183-270-2

Query Match 56.2%; Score 41; DB 4; Length 416;
Best Local Similarity 57.1%; Pred. No. 1.3e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 EWTNIWW 7
DB 220 EWSVW 226

RESULT 15
US-07-745-206A-13
Sequence 13, Application US/07745206A
Patent No. 5429921
GENERAL INFORMATION:
APPLICANT: Harpold, Michael
APPLICANT: Ellis, Steven
APPLICANT: Williams, Mark
APPLICANT: McCue, Ann
APPLICANT: Feldman, Daniel
TITLE OF INVENTION: Human Calcium Channel Compositions and
TITLE OF INVENTION: Methods
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fitch, Even, Tabin & Flannery
STREET: 135 S. LaSalle
CITY: Chicago
STATE: Illinois

COUNTRY: U.S.A.
ZIP: 60603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/745,206A
FILING DATE: 19910815
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Feder, Scott B
REFERENCE/DOCKET NUMBER: 51504
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-372-7842
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1754 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-745-206A-13

Query Match 54.8%; Score 40; DB 1; Length 1754;
Best Local Similarity 45.5%; Pred. No. 7.6e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 4; Gaps 1;
QY 2 WTNIM----WW 8
DB 1208 WTSLSWVAPW 1218

RESULT 16
US-08-311-363-13
Sequence 13, Application US/08311363
Patent No. 5876958
GENERAL INFORMATION:
APPLICANT: Harpold, Michael
APPLICANT: Ellis, Steven
APPLICANT: Williams, Mark
APPLICANT: Feldman, Daniel
APPLICANT: McCue, Ann
APPLICANT: Brenner, Robert
TITLE OF INVENTION: Human Calcium Channel Compositions and
TITLE OF INVENTION: Methods
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,363
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/745,206
FILING DATE: 15-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-51506
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 238-0999
TELEFAX: (619) 238-0062
INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:
 LENGTH: 1754 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-311-363-13

Query Match 54.8%; Score 40; DB 2; Length 1754;
 Best Local Similarity 45.5%; Pred. No. 7.6e-02;

Matches 5; Conservative 2; Mismatches 3; Indels 4; Gaps 1;

Qy 2 WTNW----WW 8
 ||:|
 Db 1208 WTSLSVAPW 1213

RESULT 17
 US-09-673-395A-570

; Sequence 570, Application US/09673395A

; Patent No. 6620923

; GENERAL INFORMATION:

; APPLICANT: SPECHT, THOMAS

; APPLICANT: HINZMANN, BERND

; APPLICANT: SCHMITT, ARMIN

; APPLICANT: PILARSKY, CHRISTIAN

; APPLICANT: DAHL, EDGAR

; APPLICANT: ROSENTHAL, ANDRE

; TITLE OF INVENTION: HUMAN NUCLEIC ACID SEQUENCES FROM UTERUS TUMOR TISSUE

; FILE REFERENCE: ALBRE-12

; CURRENT APPLICATION NUMBER: US/09/673,395A

; CURRENT FILING DATE: 2000-10-17

; NUMBER OF SEQ ID NOS: 637

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 570

; LENGTH: 199

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-673-395A-570

Query Match 54.1%; Score 39.5; DB 4; Length 199;

Best Local Similarity 55.6%; Pred. No. 95;

Matches 5; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

Qy 2 WTRINWMAK 10
 | : |||
 Db 6 W-SFWWTK 13

RESULT 18

US-09-779-451-74

; Sequence 74, Application US/09779451

; Patent No. 6605427

; GENERAL INFORMATION:

; APPLICANT: Wild, Carl T.

; APPLICANT: Allaway, Graham P.

; TITLE OF INVENTION: Assay for Detection of Viral Fusion Inhibitors

; FILE REFERENCE: 1900.030003

; CURRENT APPLICATION NUMBER: US/09/779,451

; CURRENT FILING DATE: 2001-08-17

; PRIOR APPLICATION NUMBER: US 60/235,901

; PRIOR FILING DATE: 2000-09-28

; PRIOR FILING DATE: 2000-09-28

; PRIOR FILING DATE: 2000-02-10

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 74

; LENGTH: 36

; TYPE: PRT

; ORGANISM: Human immunodeficiency virus type 1

US-09-779-451-74

Query Match

Best Local Similarity 53.4%; Score 39; DB 4; Length 36;

62.5%; Pred. No. 19;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EWTNIWW 8
 ||:|
 Db 28 EWASINW 35

RESULT 19

US-09-779-451-72

; Sequence 72, Application US/09779451

; Patent No. 6605427

; GENERAL INFORMATION:

; APPLICANT: Wild, Carl T.

; APPLICANT: Allaway, Graham P.

; TITLE OF INVENTION: Assay for Detection of Viral Fusion Inhibitors

; FILE REFERENCE: 1900.030003

; CURRENT APPLICATION NUMBER: US/09/779,451

; CURRENT FILING DATE: 2001-08-17

; PRIOR APPLICATION NUMBER: US 60/235,901

; PRIOR FILING DATE: 2000-09-28

; PRIOR APPLICATION NUMBER: US 60/181,543

; PRIOR FILING DATE: 2000-02-10

; NUMBER OF SEQ ID NOS: 77

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 72

; LENGTH: 46

; TYPE: PRT

; ORGANISM: Human immunodeficiency virus type 1

US-09-779-451-72

Query Match

Best Local Similarity 53.4%; Score 39; DB 4; Length 46;

62.5%; Pred. No. 25;

Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EWTNIWW 8
 ||:|
 Db 38 EWASINW 45

RESULT 20

US-08-630-915A-112

; Sequence 112, Application US/08630915A

; Patent No. 6309820

; GENERAL INFORMATION:

; APPLICANT: SPARKS, Andrew B.

; APPLICANT: HOFFMAN, No. 6309820h

; APPLICANT: KAY, Brian K.

; APPLICANT: FOWLES, Dana M.

; APPLICANT: MCCONNELL, Stephen J.

; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL

; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND

; TITLE OF INVENTION: USING SAME

; NUMBER OF SEQUENCES: 227

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds LLP

; STREET: 1155 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/630,915A

; FILING DATE: 03-APR-1996

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Misrock, S. Leslie

; REGISTRATION NUMBER: 18,872

; REFERENCE/DOCKET NUMBER: 1101-174

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 112:
SEQUENCE CHARACTERISTICS:
LENGTH: 60 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-915A-112

Query Match 53.4%; Score 39; DB 4; Length 60;
Best Local Similarity 80.0%; Pred. No. 33;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 WWNAK 10
Db 40 WWNAK 44

RESULT 21

US-09-252-991A-20055
Sequence 20055, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR FILING DATE: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 20055
LENGTH: 71
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20055

Query Match 53.4%; Score 39; DB 4; Length 71;
Best Local Similarity 66.7%; Pred. No. 39;
Matches 4; Conservative 1; Mismatches 0; Indels 1; Gaps 0;

QY 2 WNIWW 7
Db 54 WNTWW 59

RESULT 22

US-09-621-976-5285
Sequence 5285, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:

APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 5285
LENGTH: 92
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SIGNAL

Query Match 53.4%; Score 39; DB 4; Length 144;
Best Local Similarity 80.0%; Pred. No. 80;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 WWNAK 10
Db 113 WWNAK 117

RESULT 24

US-09-543-681A-5268
Sequence 5268, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:

LOCATION: -22...-1
US-09-621-976-5285

Query Match 53.4%; Score 39; DB 4; Length 92;
Best Local Similarity 80.0%; Pred. No. 51;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 WWNAK 10
Db 36 WWNAK 40

RESULT 23

US-08-630-915A-6
Sequence 6, Application US/08630915A
Patent No. 6309820
GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFFMAN, No. 6309820h
APPLICANT: KAY, Brian K.
APPLICANT: FOWLES, Dana M.
APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
TITLE OF INVENTION: USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/630,915A
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Wisrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 144 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-915A-6

Query Match 53.4%; Score 39; DB 4; Length 144;
Best Local Similarity 80.0%; Pred. No. 80;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 WWNAK 10
Db 113 WWNAK 117

RESULT 24

US-09-543-681A-5268
Sequence 5268, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:

APPLICANT: GARY BRYON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
PRIOR FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 5268
LENGTH: 183
TYPE: PRT
ORGANISM: Proteus mirabilis
US-09-543-681A-5268

Query Match 53.4%; Score 39; DB 4; Length 183;
Best Local Similarity 42.9%; Pred. No. 1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 6; Gaps 1;

Qy 2 WTNTW-----WVA 9
||:|
Db 168 WTSSWGLRLWVA 181

RESULT 25
US-08-311-731A-306
Sequence 306, Application US/08311731A
Patent No. 6583266
GENERAL INFORMATION:
APPLICANT: SMITH, DOUGLAS
APPLICANT: MAO, JEN-I
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
NUMBER OF SEQUENCES: 411
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/311,731A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2442
INFORMATION FOR SEQ ID NO: 306:
SEQUENCE CHARACTERISTICS:
LENGTH: 194 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Mycobacterium leprae
US-08-311-731A-306

Query Match 53.4%; Score 39; DB 4; Length 194;
Best Local Similarity 57.1%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 MTNIMWW 8
||:|
Db 55 MYPWWWW 61

RESULT 26
US-09-252-991A-19090
Sequence 19090, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107195.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 19090
LENGTH: 240
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19090

Query Match 53.4%; Score 39; DB 4; Length 240;
Best Local Similarity 44.4%; Pred. No. 1.4e+02;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 MTNIMWWAK 10
||:|
Db 226 WDSVWVWAR 234

RESULT 27
US-08-965-056-73
Sequence 73, Application US/08965056
Patent No. 6271198
GENERAL INFORMATION:
APPLICANT: Andrew C. Braisted
APPLICANT: J. Kevin Judice
APPLICANT: Robert S. McDowell
APPLICANT: J. Christopher Phelar.
APPLICANT: Melissa A. Starovasnik
APPLICANT: James A. Wells
TITLE OF INVENTION: Constrained Helical Peptides and Methods of
NUMBER OF SEQUENCES: 113
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA: US/08/965,056
FILING DATE: 05-No. 6271198-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Torchia, PhD., Timothy E.
REGISTRATION NUMBER: 36,700
REFERENCE/DOCKET NUMBER: P1005R2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-8674
TELEFAX: 650/952-9881

; INFORMATION FOR SEQ ID NO: 73:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-965-056-73

Query Match 53.4%; Score 39; DB 3; Length 243;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EWTNIWW 8
Db 195 KWTSLWSW 202

RESULT 28

US-09-479-645A-142
; Sequence 142, Application US/09479645A
; Patent No. 6489141
; GENERAL INFORMATION:
; APPLICANT: FRAZER, Ian Hector
; APPLICANT: ZHOU, Jian
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCE AND METHOD FOR SELECTIVELY
; FILE REFERENCE: 210338.0001/1US
; CURRENT APPLICATION NUMBER: US/09/479,645A
; CURRENT FILING DATE: 2000-01-07
; PRIOR APPLICATION NUMBER: PCT/AU98/00530
; PRIOR FILING DATE: 1998-07-09
; PRIOR FILING DATE: 1997-07-09
; PRIOR APPLICATION NUMBER: AU P09467
; PRIOR FILING DATE: 1997-09-11
; NUMBER OF SEQ ID NOS: 219
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 142
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Trp(TGG)5GFP
US-09-479-645A-142

Query Match 53.4%; Score 39; DB 4; Length 243;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 6 WWWAK 10
Db 4 WWWSK 8

RESULT 29

US-09-247-890-16
; Sequence 16, Application US/09247890
; Patent No. 6541011
; GENERAL INFORMATION:
; APPLICANT: Punnonen, Juha
; APPLICANT: Bass, Steven H.
; APPLICANT: Whalen, Robert Gerald
; APPLICANT: Howard, Russell
; APPLICANT: Stemmer, Willem P.C.
; APPLICANT: Maxygen, Inc.
; TITLE OF INVENTION: Antigen Library Immunization
; FILE REFERENCE: 018097-028710US
; CURRENT APPLICATION NUMBER: US/09/247,890
; CURRENT FILING DATE: 1999-02-10
; EARLIER APPLICATION NUMBER: US 60/074,294
; EARLIER FILING DATE: 1998-02-11
; EARLIER APPLICATION NUMBER: US 60/105,509
; EARLIER FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 25

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Woodchuck hepatitis B virus
US-09-247-890-16

Query Match 53.4%; Score 39; DB 4; Length 282;
Best Local Similarity 40.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

Qy 2 WTNII-----WWW 8
Db 80 WTKILTIAQNLDWW 94

RESULT 30

US-09-724-969-16
; Sequence 16, Application US/09724969
; Patent No. 6569435
; GENERAL INFORMATION:
; APPLICANT: Punnonen, Juha
; APPLICANT: Bass, Steven H.
; APPLICANT: Whalen, Robert Gerald
; APPLICANT: Howard, Russell
; APPLICANT: Stemmer, Willem P.C.
; APPLICANT: Maxygen, Inc.
; TITLE OF INVENTION: Antigen Library Immunization
; FILE REFERENCE: 018097-028710US
; CURRENT APPLICATION NUMBER: US/09/724,969
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/247,890
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/105,509
; PRIOR FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Woodchuck hepatitis B virus
US-09-724-969-16

Query Match 53.4%; Score 39; DB 4; Length 282;
Best Local Similarity 40.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

Qy 2 WTNII-----WWW 8
Db 80 WTKILTIAQNLDWW 94

RESULT 31

US-09-724-852-16
; Sequence 16, Application US/09724852
; Patent No. 6576757
; GENERAL INFORMATION:
; APPLICANT: Punnonen, Juha
; APPLICANT: Bass, Steven H.
; APPLICANT: Whalen, Robert Gerald
; APPLICANT: Howard, Russell
; APPLICANT: Stemmer, Willem P.C.
; APPLICANT: Maxygen, Inc.
; TITLE OF INVENTION: Antigen Library Immunization
; FILE REFERENCE: 018097-028710US
; CURRENT APPLICATION NUMBER: US/09/724,852
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: US/09/247,890
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/074,294
; PRIOR FILING DATE: 1998-02-11
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/105,509
; PRIOR FILING DATE: 1998-10-23

```
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Woodchuck hepatitis B virus
US-09-724-852-16

Query Match
Best Local Similarity 53.4%; Score 39; DB 4; Length 282;
Matches 6; Conservative 0; Mismatches 1; Indels 8; Gaps 1;

QY 2 WTNI-----WWW 8
DB 80 WTKILTIAQNLDWWW 94

RESULT 32
US-08-817-441-47
; Sequence 47, Application US/08817441
; Patent No. 6399294
; GENERAL INFORMATION:
; APPLICANT: CHARNEAU, PIERRE
; APPLICANT: CLAVEL, FRANCOISE
; APPLICANT: BORMAN, ANDREW
; APPLICANT: QUILLIENT, CAROLINE
; APPLICANT: GUETARD, DENISE
; APPLICANT: MONTAGNIER, LUC
; APPLICANT: DONJON DE SAINT-MARTIN, JACQUELINE
; APPLICANT: COHEN, JACQUES
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF HIV-1 TYPE (OR
; TITLE OF INVENTION: SUBTYPE) ANTIGENS
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,441
; FILING DATE: 11-JUL-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR 95/01391
; FILING DATE: 20-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9412554
; FILING DATE: 20-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9502526
; FILING DATE: 03-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03260.6005-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-817-441-47

Query Match
Best Local Similarity 53.4%; Score 39; DB 4; Length 862;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNIWWW 8
DB 153 EWASINWW 160

RESULT 33
US-09-462-917A-134
; Sequence 134, Application US/09462917A
; Patent No. 6511801
; GENERAL INFORMATION:
; APPLICANT: Delaporte, Eric
; APPLICANT: Peeters, Martine
; APPLICANT: Saman, Eric
; APPLICANT: Vanden Haesevelde, Marlen
; TITLE OF INVENTION: HIV-1 Group O Antigens and Uses Thereof
; FILE REFERENCE: INNS:014 11362.0014.NPUS00
; CURRENT APPLICATION NUMBER: US/09/462,917A
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: PCT/EP98/04522
; PRIOR FILING DATE: 1998-07-20
; NUMBER OF SEQ ID NOS: 152
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 134
; LENGTH: 715
; TYPE: PRT
; ORGANISM: Human
US-09-462-917A-134

Query Match
Best Local Similarity 53.4%; Score 39; DB 4; Length 715;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNIWWW 8
DB 702 EWASINWW 709

RESULT 34
US-09-206-551-15
; Sequence 15, Application US/09206551B
; Patent No. 6521739
; GENERAL INFORMATION:
; APPLICANT: Hahn, Beatrice H.
; APPLICANT: Gao, Feng
; APPLICANT: Marx, Preston A.
; APPLICANT: Shaw, George M.
; APPLICANT: Smith, Stephen M.
; APPLICANT: Georges-Courbot, Marie Claude
; APPLICANT: Lu, Chang Yong
; TITLE OF INVENTION: Complete Genome Sequences of a Simian
; TITLE OF INVENTION: Immunodeficiency Virus from a Red-Capped
; FILE REFERENCE: D6286
; CURRENT APPLICATION NUMBER: US/09/206,551B
; CURRENT FILING DATE: 1998-12-07
; NUMBER OF SEQ ID NOS: 58
; SEQ ID NO 15
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Simian immunodeficiency virus
; FEATURES:
; OTHER INFORMATION: Amino acid sequence of homologous region of
; OTHER INFORMATION: O_ANT70 lentiviral env protein
US-09-206-551-15

Query Match
Best Local Similarity 53.4%; Score 39; DB 4; Length 862;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```

Qy 1 EWTNIWNW 8
|| :|||
Db 664 EWASINW 671

RESULT 35

US-08-817-441-102
; Sequence 102, Application US/08817441
; Patent No. 6399294
; GENERAL INFORMATION:
; APPLICANT: CHARNEAU, PIERRE
; APPLICANT: CLAVEL, FRANCOISE
; APPLICANT: BORMAN, ANDREW
; APPLICANT: QUILLIENT, CAROLINE
; APPLICANT: GUETARD, DENISE
; APPLICANT: MONTAGNIER, LUC
; APPLICANT: DONON DE SAINT-MARTIN, JACQUELINE
; APPLICANT: COHEN, JACQUES
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF HIV-1 TYPE (OR
; TITLE OF INVENTION: SUBTYPE) ANTIGENS
; NUMBER OF SEQUENCES: 103
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Parabow, Garrett &
; ADDRESSEE: Dunner, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,441
; FILING DATE: 11-JUL-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR 95/01391
; FILING DATE: 20-OCT-1995
; PRIOR APPLICATION DATA: FR 9412554
; FILING DATE: 20-OCT-1994
; APPLICATION NUMBER: FR 9502526
; FILING DATE: 03-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03260.6005-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 877 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-817-441-102

Query Match 53.4%; Score 39; DB 4; Length 877;
Best Local Similarity 62.5%; Pred. No. 5.1e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EWTNIWNW 8
|| :|||
Db 679 EWASINW 686

RESULT 36

US-08-617-785-2
; Sequence 2, Application US/08617785E
; Patent No. 6228610
; GENERAL INFORMATION:
; APPLICANT: Flor, Peter J.
; APPLICANT: Kuhn, Ranier
; APPLICANT: Lindaur, Kristen
; APPLICANT: Puttner, Irene
; APPLICANT: Knopf, Thomas
; TITLE OF INVENTION: Human Metabotropic Glutamate Receptor Subtypes (HMR4,
; FILE REFERENCE: HMR6, HMR7) and Related DNA Compounds
; FILE REFERENCE: 4-19679/A/PCT
; CURRENT APPLICATION NUMBER: US/08/617,785E
; CURRENT FILING DATE: 1996-03-19
; EARLIER APPLICATION NUMBER: PCT/EP94/02991
; EARLIER FILING DATE: 1994-09-07
; EARLIER APPLICATION NUMBER: EPO 9416553.7
; EARLIER FILING DATE: 1994-08-19
; EARLIER APPLICATION NUMBER: EPO 93810663.0
; EARLIER FILING DATE: 1993-09-20
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 912
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-617-785-2

Query Match 53.4%; Score 39; DB 3; Length 912;
Best Local Similarity 80.0%; Pred. No. 5.3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 6 WWMK 10
|||||
Db 9 WWMK 13

RESULT 37

US-09-641-318-2
; Sequence 2, Application US/09641318
; Patent No. 6384205
; GENERAL INFORMATION:
; APPLICANT: BELAGAJE, RAMA M.
; WU, SU
; TITLE OF INVENTION: EXCITATORY AMINO ACID RECEPTOR PROTEIN
; AND RELATED NUCLEIC ACID COMPOUNDS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ELI LILLY AND COMPANY
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: INDIANA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/641,318
; FILING DATE: 18-AUG-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/816,178A
; FILING DATE: 12-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: GAYLO, PAUL J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-10579
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 2:

```
;
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 912 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-641-318-2

Query Match          53.4%; Score 39; DB 4; Length 912;
Best Local Similarity 80.0%; Pred. No. 5.3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      6 WWWAK 10
Db      9 WWWAR 13

RESULT 38
US-09-817-464-2
; Sequence 2, Application US/09817464
; Patent No. 6515107
; GENERAL INFORMATION:
; APPLICANT: Flor, Peter J.
; APPLICANT: Kuhn, Ranier
; APPLICANT: Lindaur, Kristen
; APPLICANT: Puttner, Irene
; APPLICANT: Knopfel, Thomas
; TITLE OF INVENTION: Human Metabotropic Glutamate Receptor Subtypes (HMR4,
; TITLE OF INVENTION: HMR6, HMR7) and Related DNA Compounds
; FILE REFERENCE: 4-19679/A/PCT
; CURRENT APPLICATION NUMBER: US/09/817,464
; CURRENT FILING DATE: 2001-03-26
; EARLIER APPLICATION NUMBER: US/08/617,785
; EARLIER FILING DATE: 1996-03-19
; EARLIER APPLICATION NUMBER: EPO 9416553.7
; EARLIER FILING DATE: 1994-08-19
; EARLIER APPLICATION NUMBER: EPO 93810663.0
; EARLIER FILING DATE: 1993-09-20
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 912
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-817-464-2

Query Match          53.4%; Score 39; DB 4; Length 912;
Best Local Similarity 80.0%; Pred. No. 5.3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      6 WWWAK 10
Db      9 WWWAR 13

RESULT 39
PCT-US91-09422-19
; Sequence 19, Application PC/TUS9109422
; GENERAL INFORMATION:
; APPLICANT: Mulvihill, Eileen R.
; APPLICANT: Hagen, Frederick S.
; APPLICANT: Houamed, Khaled M.
; APPLICANT: Almers, Wolfhard
; TITLE OF INVENTION: G PROTEIN-COUPLED GLUTAMATE RECEPTORS
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend
; STREET: One Market Plaza, Stewart Street Tower
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105-1492
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/09422
; FILING DATE: 19911212
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,307
; FILING DATE: 18-MAR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/648,481
; FILING DATE: 30-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/626,806
; FILING DATE: 12-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W.
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 13952-6PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 467-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 912 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US91-09422-19

Query Match          53.4%; Score 39; DB 5; Length 912;
Best Local Similarity 80.0%; Pred. No. 5.3e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      6 WWWAK 10
Db      9 WWWAR 13

RESULT 40
US-08-935-450-2
; Sequence 2, Application US/08935450
; Patent No. 5977311
; GENERAL INFORMATION:
; APPLICANT: Mandabalan, Krishnan
; APPLICANT: Yang, Meijia
; APPLICANT: Schulz, Vincent
; TITLE OF INVENTION: 53BP2 COMPLEXES
; FILE REFERENCE: 7934-054
; CURRENT APPLICATION NUMBER: US/08/935,450
; CURRENT FILING DATE: 1997-09-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1005
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-935-450-2

Query Match          53.4%; Score 39; DB 2; Length 1005;
Best Local Similarity 80.0%; Pred. No. 5.9e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      6 WWWAK 10
Db      974 WWWAR 978

RESULT 41
US-09-338-123-2
; Sequence 2, Application US/09338123
```

Patent No. 6627405
GENERAL INFORMATION:
APPLICANT: Nandabalan, Krishnan
APPLICANT: Yang, MeiJia
APPLICANT: Schulz, Vincent
TITLE OF INVENTION: 53BP2 COMPLEXES
FILE REFERENCE: 7934-054
CURRENT APPLICATION NUMBER: US/09/338,123
CURRENT FILING DATE: 1999-06-22
EARLIER APPLICATION NUMBER: 08/935,450
EARLIER FILING DATE: 1997-09-23
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 1005
TYPE: PRT
ORGANISM: Homo sapiens
US-09-338-123-2

Query Match 53.4%; Score 39; DB 4; Length 1005;
Best Local Similarity 80.0%; Pred. No. 5.9e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 WNAK 10
DB 974 WNAK 978

RESULT 42
US-08-825-852-49
Sequence 49, Application US/08825852
Patent No. 6121416
GENERAL INFORMATION:
APPLICANT: Clark, Ross G1
APPLICANT: Lowman, Henry B.
APPLICANT: Robinson, Iain C.A.F.
TITLE OF INVENTION: Insulin-like Growth Factor Agonist Molecules
NUMBER OF SEQUENCES: 79
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/825,852
FILING DATE: 04-Apr-1997
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: P1071
TELEPHONE: 650/225-1896
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-825-852-49

Query Match 52.7%; Score 38.5; DB 3; Length 20;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 5; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 EWTNIN-WW 8
DB 1 EWTNIN-WW 8

RESULT 44
US-09-723-890-50
Sequence 50, Application US/09723890
Patent No. 6608031
GENERAL INFORMATION:
APPLICANT: Clark, Ross G1
APPLICANT: Lowman, Henry B.
APPLICANT: Robinson, Iain C.A.F.
TITLE OF INVENTION: Insulin-like Growth Factor Agonist Molecules
NUMBER OF SEQUENCES: 109
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible

Db 10 EWAWIWW 18

RESULT 43
US-09-052-888-50
Sequence 50, Application US/09052888
Patent No. 6251865
GENERAL INFORMATION:
APPLICANT: Clark, Ross G1
APPLICANT: Lowman, Henry B.
APPLICANT: Robinson, Iain C.A.F.
TITLE OF INVENTION: Insulin-like Growth Factor Agonist Molecules
NUMBER OF SEQUENCES: 109
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/052,888
FILING DATE: 31-Mar-1998
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: P1071P1
TELEPHONE: 650/225-1896
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-09-052-888-50

Query Match 52.7%; Score 38.5; DB 3; Length 20;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 EWTNIN-WW 8
DB 10 EWAWIWW 18

RESULT 44
US-09-723-890-50
Sequence 50, Application US/09723890
Patent No. 6608031
GENERAL INFORMATION:
APPLICANT: Clark, Ross G1
APPLICANT: Lowman, Henry B.
APPLICANT: Robinson, Iain C.A.F.
TITLE OF INVENTION: Insulin-like Growth Factor Agonist Molecules
NUMBER OF SEQUENCES: 109
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible

```
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/723,890
; FILING DATE: 28-Mar-1998
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/052,888
; FILING DATE: 31-Mar-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: P1071P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1896
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 50:
;
US-09-723-890-50
;
Query Match 52.7%; Score 38.5; DB 4; Length 20;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 EWTNIM-WW 8
DB 10 EWAWIWEWW 18

RESULT 45
US-09-723-901-50
; Sequence 50, Application US/09723901
; Patent No. 6620789
; GENERAL INFORMATION:
; APPLICANT: Clark, Ross G1
; Lowman, Henry B.
; Robinson, Iain C.A.F.
; TITLE OF INVENTION: Insulin-like Growth Factor Agonist Molecules
; NUMBER OF SEQUENCES: 109
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/723,901
; FILING DATE: 28-Mar-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/052,888
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: P1071P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1896
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 50:
;
US-09-723-901-50
;
Query Match 52.7%; Score 38.5; DB 4; Length 20;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 EWTNIM-WW 8
DB 10 EWAWIWEWW 18

RESULT 46
US-09-723-547-50
; Sequence 50, Application US/09723547
; Patent No. 6632794
; GENERAL INFORMATION:
; APPLICANT: Clark, Ross G1
; Lowman, Henry B.
; Robinson, Iain C.A.F.
; TITLE OF INVENTION: Insulin-like Growth Factor Agonist Molecules
; NUMBER OF SEQUENCES: 109
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/723,547
; FILING DATE: 28-Mar-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/052,888
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: P1071P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1896
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 50:
;
US-09-723-547-50
;
Query Match 52.7%; Score 38.5; DB 4; Length 20;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 EWTNIM-WW 8
DB 10 EWAWIWEWW 18

RESULT 47
US-09-724-127-50
; Sequence 50, Application US/09724127
; Patent No. 6635619
; GENERAL INFORMATION:
; APPLICANT: Clark, Ross G1
; Lowman, Henry B.
```

```
;
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 50:
;
US-09-723-901-50
;
Query Match 52.7%; Score 38.5; DB 4; Length 20;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 EWTNIM-WW 8
DB 10 EWAWIWEWW 18

RESULT 46
US-09-723-547-50
; Sequence 50, Application US/09723547
; Patent No. 6632794
; GENERAL INFORMATION:
; APPLICANT: Clark, Ross G1
; Lowman, Henry B.
; Robinson, Iain C.A.F.
; TITLE OF INVENTION: Insulin-like Growth Factor Agonist Molecules
; NUMBER OF SEQUENCES: 109
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/723,547
; FILING DATE: 28-Mar-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/052,888
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: P1071P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1896
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 50:
;
US-09-723-547-50
;
Query Match 52.7%; Score 38.5; DB 4; Length 20;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 EWTNIM-WW 8
DB 10 EWAWIWEWW 18

RESULT 47
US-09-724-127-50
; Sequence 50, Application US/09724127
; Patent No. 6635619
; GENERAL INFORMATION:
; APPLICANT: Clark, Ross G1
; Lowman, Henry B.
```


Qy 1 EWTN1W-WW 8
Db 10 EWAW1EW 18

RESULT 5C
US-09-724-114-50
; Sequence 50, Application US/09724114
; Patent No. 6680298
; GENERAL INFORMATION:
; APPLICANT: Clark, Ross G1
; Lowman, Henry B.
; Robinson, Iain C.A.F.
; TITLE OF INVENTION: Insulin-like Growth Factor Agonist Molecules
; NUMBER OF SEQUENCES: 109
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94030
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/724,114
; FILING DATE: 28-Mar-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/052,888
; FILING DATE: 31-Mar-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: P1071P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1896
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-09-724-114-50

Query Match 52.7%; Score 38.5; DB 4; Length 20;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 1 EWTN1W-WW 8
Db 10 EWAW1EW 18

RESULT 51
US-09-723-913-50
; Sequence 50, Application US/09723913
; Patent No. 6683053
; GENERAL INFORMATION:
; APPLICANT: Clark, Ross G1
; Lowman, Henry B.
; Robinson, Iain C.A.F.
; TITLE OF INVENTION: Insulin-like Growth Factor Agonist Molecules
; NUMBER OF SEQUENCES: 109
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California

COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/723,913
FILING DATE: 28-Mar-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/052,888
FILING DATE: 31-Mar-1998
ATTORNEY/AGENT INFORMATION:
NAME: Hasak, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: P1071P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-1896
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-09-723-913-50

Query Match 52.7%; Score 38.5; DB 4; Length 20;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 1 EWTN1W-WW 8
Db 10 EWAW1EW 18

RESULT 52
US-09-424-349A-2
; Sequence 2, Application US/09424349A
; Patent No. 6387668
; GENERAL INFORMATION:
; APPLICANT: Lucje Spelberg, Jeffrey Harald
; Rink, Rick
; Kellogg, Richard Morrison
; Janssen, Dick Barend
; TITLE OF INVENTION: Enantioselective epoxide hydrolases and genes
; encoding these.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Vereenigde
; STREET: Nieuwe Parklaan 97
; CITY: The Hague
; STATE: Zuid-Holland
; COUNTRY: The Netherlands
; ZIP: 2587 BN
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/424,349A
FILING DATE: 23-Feb-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 97201515.0
FILING DATE: 21-MAY-1997
APPLICATION NUMBER: PCT/NL98/00290
FILING DATE: 20-MAY-1998
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 294 amino acids


```
;
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-424-349A-2

Query Match          52.7%; Score 38.5; DB 4; Length 294;
Best Local Similarity 50.0%; Pred. No. 2e+02;
Matches 5; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Qy 2 WTNIWV-WAK 10
Db 38 WPGFWWWSK 47

RESULT 53
US-09-424-349A-9
; Sequence 9, Application US/09424349A
; Patent No. 6387668
; GENERAL INFORMATION:
; APPLICANT: Lütje Spelberg, Jeffrey Harald
; Rink, Rick
; Kellogg, Richard Morrison
; Janssen, Dick Barend
; TITLE OF INVENTION: Enantioselective epoxide hydrolases and genes
; encoding these.
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Vereenigde
; STREET: Nieuwe Parklaan 97
; CITY: The Hague
; STATE: Zuid-Holland
; COUNTRY: The Netherlands
; ZIP: 2587 BN
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/424,349A
; FILING DATE: 23-Feb-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 97201515.0
; FILING DATE: 21-MAY-1997
; APPLICATION NUMBER: PCT/NL98/00290
; FILING DATE: 20-MAY-1998
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 294 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-424-349A-3

Query Match          52.7%; Score 38.5; DB 4; Length 294;
Best Local Similarity 50.0%; Pred. No. 2e+02;
Matches 5; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Qy 2 WTNIWV-WAK 10
Db 38 WPGFWWWSK 47

RESULT 54
US-09-082-279B-1499
; Sequence 1499, Application US/09082279B
; Patent No. 6258782
```

```
; GENERAL INFORMATION:
; APPLICANT: Barney, Shawn
; APPLICANT: Guthrie, Kelly
; APPLICANT: Merutka, Gene
; APPLICANT: Anwer, Mohamed
; APPLICANT: Lambert, Dennis
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED
; PHARMACOKINETIC PROPERTIES
; FILE REFERENCE: 7872-043
; CURRENT APPLICATION NUMBER: US/09/082,279B
; CURRENT FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1515
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1499
; LENGTH: 8
; TYPE: PRT
; ORGANISM: HIV-1
US-09-082-279B-1499
```

```
Query Match          52.1%; Score 38; DB 3; Length 8;
Best Local Similarity 57.1%; Pred. No. 3e+05;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 WTNIWVW 8
Db 1 WTSLSW 7
```

```
RESULT 55
US-09-082-279B-1507
; Sequence 1507, Application US/09082279B
; Patent No. 6258782
; GENERAL INFORMATION:
; APPLICANT: Barney, Shawn
; APPLICANT: Guthrie, Kelly
; APPLICANT: Merutka, Gene
; APPLICANT: Anwer, Mohamed
; APPLICANT: Lambert, Dennis
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED
; PHARMACOKINETIC PROPERTIES
; FILE REFERENCE: 7872-043
; CURRENT APPLICATION NUMBER: US/09/082,279B
; CURRENT FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1515
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1507
; LENGTH: 8
; TYPE: PRT
; ORGANISM: HIV-1
US-09-082-279B-1507
```

```
Query Match          52.1%; Score 38; DB 3; Length 8;
Best Local Similarity 57.1%; Pred. No. 3e+05;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 WTNIWVW 8
Db 1 WANLSW 7
```

```
RESULT 56
US-09-315-304B-1653
; Sequence 1653, Application US/09315304B
; Patent No. 6348568
; GENERAL INFORMATION:
; APPLICANT: Barney, S.
; APPLICANT: Guthrie, K.
; APPLICANT: Merutka, G.
; APPLICANT: Anwer, M.
; APPLICANT: Lambert, D.
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED PHARMACOKINETIC
; PROPERTIES
; FILE REFERENCE: 7872-052
```

; CURRENT APPLICATION NUMBER: US/09/315,304B
; CURRENT FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 09/082,279
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1667
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1653
; LENGTH: 8
; TYPE: PRT
; ORGANISM: HIV-1
US-09-315-304B-1653

Query Match 52.1%; Score 38; DB 4; Length 8;
Best Local Similarity 57.1%; Pred. No. 3e+05;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WTNWVW 8
Db 1 WWSLWSW 7

RESULT 57
US-09-834-784-1499
; Sequence 1499, Application US/09834784
; Patent No. 6562787
; GENERAL INFORMATION:
; APPLICANT: Barney, Shawn
; APPLICANT: Guthrie, Kelly
; APPLICANT: Merutka, Gene
; APPLICANT: Anwer, Mohamed
; APPLICANT: Lambert, Dennis
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED
; FILE REFERENCE: 7872-043
; CURRENT APPLICATION NUMBER: US/09/834,784
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 09/082,279
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1515
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1499
; LENGTH: 8
; TYPE: PRT
; ORGANISM: HIV-1
US-09-834-784-1499

Query Match 52.1%; Score 38; DB 4; Length 8;
Best Local Similarity 57.1%; Pred. No. 3e+05;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WTNWVW 8
Db 1 WWSLWSW 7

RESULT 58
US-09-834-784-1507
; Sequence 1507, Application US/09834784
; Patent No. 6562787
; GENERAL INFORMATION:
; APPLICANT: Barney, Shawn
; APPLICANT: Guthrie, Kelly
; APPLICANT: Merutka, Gene
; APPLICANT: Anwer, Mohamed
; APPLICANT: Lambert, Dennis
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED
; FILE REFERENCE: 7872-043
; CURRENT APPLICATION NUMBER: US/09/834,784
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 09/082,279
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1515

; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1507
; LENGTH: 8
; TYPE: PRT
; ORGANISM: HIV-1
US-09-834-784-1507

Query Match 52.1%; Score 38; DB 4; Length 8;
Best Local Similarity 57.1%; Pred. No. 3e+05;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 WTNWVW 8
Db 1 WWSLWSW 7

RESULT 59
US-09-350-641C-1654
; Sequence 1654, Application US/09350641C
; Patent No. 6656906
; GENERAL INFORMATION:
; APPLICANT: Barney, S.
; APPLICANT: Guthrie, K.
; APPLICANT: Merutka, G.
; APPLICANT: Anwer, M.
; APPLICANT: Lambert, D.
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED PHARMACOKINETIC
; FILE REFERENCE: 7872-067
; CURRENT APPLICATION NUMBER: US/09/350,641C
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/315,304
; PRIOR FILING DATE: 1998-05-20
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1757
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1654
; LENGTH: 8
; TYPE: PRT
; ORGANISM: HIV-1
US-09-350-641C-1654

Query Match 52.1%; Score 38; DB 4; Length 8;
Best Local Similarity 57.1%; Pred. No. 3e+05;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WTNWVW 8
Db 1 WWSLWSW 7

RESULT 60
US-09-350-641C-1662
; Sequence 1662, Application US/09350641C
; Patent No. 6656906
; GENERAL INFORMATION:
; APPLICANT: Barney, S.
; APPLICANT: Guthrie, K.
; APPLICANT: Merutka, G.
; APPLICANT: Anwer, M.
; APPLICANT: Lambert, D.
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED PHARMACOKINETIC
; FILE REFERENCE: 7872-067
; CURRENT APPLICATION NUMBER: US/09/350,641C
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/315,304
; PRIOR FILING DATE: 1999-05-20
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1757
; SOFTWARE: FastSEQ for Windows Version 3.0

```
; SEQ ID NO 1662
; LENGTH: 8
; TYPE: PRT
; ORGANISM: HIV-1
US-09-350-641C-1662

Query Match      52.1%; Score 38; DB 4; Length 8;
Best Local Similarity 57.1%; Pred. No. 3e+05;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 EWTNIWW 8
   ||::||
Db 1 WANLWSW 7

RESULT 61
US-09-315-304B-1564
; Sequence 1564, Application US/09315304B
; Patent No. 6348568
; GENERAL INFORMATION:
; APPLICANT: Barney, S.
; APPLICANT: Guthrie, K.
; APPLICANT: Merutka, G.
; APPLICANT: Anwer, M.
; APPLICANT: Lambert, D.
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED PHARMACOKINETIC
; TITLE OF INVENTION: PROPERTIES
; FILE REFERENCE: 7872-052
; CURRENT APPLICATION NUMBER: US/09/315,304B
; CURRENT FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 09/082,279
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1667
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1564
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Core polypeptide
US-09-315-304B-1564

Query Match      52.1%; Score 38; DB 4; Length 9;
Best Local Similarity 50.3%; Pred. No. 3e+05;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNIWW 8
   ||::||
Db 1 EWASLW 8

RESULT 62
US-09-315-304B-1580
; Sequence 1580, Application US/09315304B
; Patent No. 6348568
; GENERAL INFORMATION:
; APPLICANT: Barney, S.
; APPLICANT: Guthrie, K.
; APPLICANT: Merutka, G.
; APPLICANT: Anwer, M.
; APPLICANT: Lambert, D.
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED PHARMACOKINETIC
; TITLE OF INVENTION: PROPERTIES
; FILE REFERENCE: 7872-052
; CURRENT APPLICATION NUMBER: US/09/315,304B
; CURRENT FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 09/082,279
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1667
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1580
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Core polypeptide
US-09-315-304B-1580

Query Match      52.1%; Score 38; DB 4; Length 9;
Best Local Similarity 50.0%; Pred. No. 3e+05;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNIWW 8
   ||::||
Db 1 EWASLW 8

RESULT 63
US-09-350-325-24
; Sequence 24, Application US/09350325
; Patent No. 6541020
; GENERAL INFORMATION:
; APPLICANT: Ding, S.
; APPLICANT: Kang, M.
; APPLICANT: Venetta, T.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ADMINISTRATION OF
; TITLE OF INVENTION: THERAPEUTIC REAGENTS
; FILE REFERENCE: 7872-062
; CURRENT APPLICATION NUMBER: US/09/350,325
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: enhancer peptide
US-09-350-325-24

Query Match      52.1%; Score 38; DB 4; Length 9;
Best Local Similarity 50.0%; Pred. No. 3e+05;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNIWW 8
   ||::||
Db 1 EWASLW 8

RESULT 64
US-09-350-325-40
; Sequence 40, Application US/09350325
; Patent No. 6541020
; GENERAL INFORMATION:
; APPLICANT: Ding, S.
; APPLICANT: Kang, M.
; APPLICANT: Venetta, T.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ADMINISTRATION OF
; TITLE OF INVENTION: THERAPEUTIC REAGENTS
; FILE REFERENCE: 7872-062
; CURRENT APPLICATION NUMBER: US/09/350,325
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 40
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: enhancer peptide
US-09-350-325-40

Query Match      52.1%; Score 38; DB 4; Length 9;
Best Local Similarity 50.0%; Pred. No. 3e+05;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 EWTNIWW 8
   || :||
Db 2 EWASLW 9

Best Local Similarity 50.0%; Pred. No. 3e+05;
Matches 4; Conservative 2; Mismatches 0; Indels 2; Gaps 0;

RESULT 65
US-09-350-641C-1564
; Sequence 1564, Application US/09350641C
; Patent No. 6656906
; GENERAL INFORMATION:
; APPLICANT: Barney, S.
; APPLICANT: Guthrie, K.
; APPLICANT: Merutka, G.
; APPLICANT: Anwer, M.
; APPLICANT: Lambert, D.
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED PHARMACOKINETIC
; TITLE OF INVENTION: PROPERTIES
; FILE REFERENCE: 7872-067
; CURRENT APPLICATION NUMBER: US/09/350,641C
; CURRENT FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/315,304
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 09/082,279
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1757
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1564
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Core polypeptide
US-09-350-641C-1564

Query Match 52.1%; Score 38; DB 4; Length 9;
Best Local Similarity 50.0%; Pred. No. 3e+05;
Matches 4; Conservative 2; Mismatches 0; Indels 2; Gaps 0;

QY 1 EWTNIWW 8
   || :||
Db 1 EWASLW 9

RESULT 66
US-09-350-641C-1580
; Sequence 1580, Application US/09350641C
; Patent No. 6656906
; GENERAL INFORMATION:
; APPLICANT: Barney, S.
; APPLICANT: Guthrie, K.
; APPLICANT: Merutka, G.
; APPLICANT: Anwer, M.
; APPLICANT: Lambert, D.
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED PHARMACOKINETIC
; TITLE OF INVENTION: PROPERTIES
; FILE REFERENCE: 7872-067
; CURRENT APPLICATION NUMBER: US/09/350,641C
; CURRENT FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/315,304
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 09/082,279
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1757
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1580
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Core polypeptide
US-09-350-641C-1580

Query Match 52.1%; Score 38; DB 4; Length 9;
Best Local Similarity 50.0%; Pred. No. 3e+05;
Matches 4; Conservative 2; Mismatches 0; Indels 2; Gaps 0;

QY 1 EWTNIWW 8
   || :||
Db 1 EWASLW 9

RESULT 67
US-09-315-304B-1573
; Sequence 1573, Application US/09315304B
; Patent No. 6348568
; GENERAL INFORMATION:
; APPLICANT: Barney, S.
; APPLICANT: Guthrie, K.
; APPLICANT: Merutka, G.
; APPLICANT: Anwer, M.
; APPLICANT: Lambert, D.
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED PHARMACOKINETIC
; TITLE OF INVENTION: PROPERTIES
; FILE REFERENCE: 7872-052
; CURRENT APPLICATION NUMBER: US/09/315,304B
; CURRENT FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 09/082,279
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1667
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1573
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Core polypeptide
US-09-315-304B-1573

Query Match 52.1%; Score 38; DB 4; Length 10;
Best Local Similarity 50.0%; Pred. No. 7.2;
Matches 4; Conservative 2; Mismatches 2; Indels 3; Gaps 0;

QY 1 EWTNIWW 8
   || :||
Db 2 EWASLW 9

RESULT 68
US-09-315-304B-1575
; Sequence 1575, Application US/09315304B
; Patent No. 6348568
; GENERAL INFORMATION:
; APPLICANT: Barney, S.
; APPLICANT: Guthrie, K.
; APPLICANT: Merutka, G.
; APPLICANT: Anwer, M.
; APPLICANT: Lambert, D.
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED PHARMACOKINETIC
; TITLE OF INVENTION: PROPERTIES
; FILE REFERENCE: 7872-052
; CURRENT APPLICATION NUMBER: US/09/315,304B
; CURRENT FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 09/082,279
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1667
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1575
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Core polypeptide
US-09-315-304B-1575

Query Match 52.1%; Score 38; DB 4; Length 10;
Best Local Similarity 50.0%; Pred. No. 7.2;
Matches 4; Conservative 2; Mismatches 2; Indels 3; Gaps 0;
```

Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EWTNIWW 8
|| :||
Db 2 EWASLW 9

RESULT 69

US-09-315-304B-1584
; Sequence 1584, Application US/09315304B
; Patent No. 6348568
; GENERAL INFORMATION:
; APPLICANT: Barney, S.
; APPLICANT: Guthrie, K.
; APPLICANT: Merutka, G.
; APPLICANT: Anwer, M.
; APPLICANT: Lambert, D.
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED PHARMACOKINETIC
; TITLE OF INVENTION: PROPERTIES
; FILE REFERENCE: 7872-052
; CURRENT APPLICATION NUMBER: US/09/315.304B
; CURRENT FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 09/082,279
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1667
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1584
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Core polypeptide
US-09-315-304B-1584

Query Match 52.1%; Score 38; DB 4; Length 10;
Best Local Similarity 50.0%; Pred. No. 7.2;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EWTNIWW 8
|| :||
Db 2 EWASLW 9

RESULT 70

US-09-350-325-33
; Sequence 33, Application US/09350325
; Patent No. 6541020
; GENERAL INFORMATION:
; APPLICANT: Ding, S.
; APPLICANT: Kang, M.
; APPLICANT: Venetta, T.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ADMINISTRATION OF
; TITLE OF INVENTION: THERAPEUTIC REAGENTS
; FILE REFERENCE: 7872-062
; CURRENT APPLICATION NUMBER: US/09/350,325
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 33
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: enhancer peptide
US-09-350-325-33

Query Match 52.1%; Score 38; DB 4; Length 10;
Best Local Similarity 50.0%; Pred. No. 7.2;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EWTNIWW 8
|| :||
Db 2 EWASLW 9

RESULT 71

US-09-350-325-35
; Sequence 35, Application US/09350325
; Patent No. 6541020
; GENERAL INFORMATION:
; APPLICANT: Ding, S.
; APPLICANT: Kang, M.
; APPLICANT: Venetta, T.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ADMINISTRATION OF
; TITLE OF INVENTION: THERAPEUTIC REAGENTS
; FILE REFERENCE: 7872-062
; CURRENT APPLICATION NUMBER: US/09/350,325
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 35
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: enhancer peptide
US-09-350-325-35

Query Match 52.1%; Score 38; DB 4; Length 10;
Best Local Similarity 50.0%; Pred. No. 7.2;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EWTNIWW 8
|| :||
Db 2 EWASLW 9

RESULT 72

US-09-350-325-44
; Sequence 44, Application US/09350325
; Patent No. 6541020
; GENERAL INFORMATION:
; APPLICANT: Ding, S.
; APPLICANT: Kang, M.
; APPLICANT: Venetta, T.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ADMINISTRATION OF
; TITLE OF INVENTION: THERAPEUTIC REAGENTS
; FILE REFERENCE: 7872-062
; CURRENT APPLICATION NUMBER: US/09/350,325
; CURRENT FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 44
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: enhancer peptide
US-09-350-325-44

Query Match 52.1%; Score 38; DB 4; Length 10;
Best Local Similarity 50.0%; Pred. No. 7.2;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 EWTNIWW 8
|| :||
Db 2 EWASLW 9

RESULT 73

US-09-350-641C-1573
; Sequence 1573, Application US/09350641C
; Patent No. 6656906
; GENERAL INFORMATION:
; APPLICANT: Barney, S.
; APPLICANT: Guthrie, K.

```
; APPLICANT: Merutka, G.
; APPLICANT: Anwer, M.
; APPLICANT: Lambert, D.
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED PHARMACOKINETIC
; TITLE OF INVENTION: PROPERTIES
; FILE REFERENCE: 7872-067
; CURRENT APPLICATION NUMBER: US/09/350,641C
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/315,304
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 09/082,279
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1757
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1573
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Core polypeptide
US-09-350-641C-1573
```

```
Query Match 52.1%; Score 38; DB 4; Length 10;
Best Local Similarity 50.0%; Pred. No. 7.2;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 EWTNIWW 8
Db 2 EWASLW 9
```

RESULT 74

```
US-09-350-641C-1575
; Sequence 1575, Application US/09350641C
; Patent No. 6656906
; GENERAL INFORMATION:
; APPLICANT: Barney, S.
; APPLICANT: Guthrie, K.
; APPLICANT: Merutka, G.
; APPLICANT: Anwer, M.
; APPLICANT: Lambert, D.
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED PHARMACOKINETIC
; TITLE OF INVENTION: PROPERTIES
; FILE REFERENCE: 7872-067
; CURRENT APPLICATION NUMBER: US/09/350,641C
; CURRENT FILING DATE: 1999-07-09
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 09/315,304
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1757
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1575
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Core polypeptide
US-09-350-641C-1575
```

```
Query Match 52.1%; Score 38; DB 4; Length 10;
Best Local Similarity 50.0%; Pred. No. 7.2;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 EWTNIWW 8
Db 2 EWASLW 9
```

RESULT 75

```
US-09-350-641C-1584
; Sequence 1584, Application US/09350641C
; Patent No. 6656906
```

```
; GENERAL INFORMATION:
; APPLICANT: Barney, S.
; APPLICANT: Guthrie, K.
; APPLICANT: Merutka, G.
; APPLICANT: Anwer, M.
; APPLICANT: Lambert, D.
; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED PHARMACOKINETIC
; TITLE OF INVENTION: PROPERTIES
; FILE REFERENCE: 7872-067
; CURRENT APPLICATION NUMBER: US/09/350,641C
; CURRENT FILING DATE: 1999-07-09
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 09/315,304
; PRIOR FILING DATE: 1998-05-20
; PRIOR APPLICATION NUMBER: 09/082,279
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1757
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1584
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Core polypeptide
US-09-350-641C-1584
```

```
Query Match 52.1%; Score 38; DB 4; Length 10;
Best Local Similarity 50.0%; Pred. No. 7.2;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 EWTNIWW 8
Db 2 EWASLW 9
```

```
Search completed: June 9, 2004, 18:12:06
Job time : 34 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2004, 18:11:23 ; Search time 42 Seconds
(without alignments)
66.985 Million cell updates/sec

Title: US-10-726-148A-15_COPY_428_437

Perfect score: 73

Sequence: 1 EWTNIWMAX 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1:55919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	72.6	53	12	US-10-424-599-152002, Sequence 152002,
2	45	61.6	88	12	US-10-424-599-276989, Sequence 276989,
3	45	61.6	99	12	US-10-425-114-53192, Sequence 53192, A
4	45	61.6	189	12	US-10-425-114-47464, Sequence 47464, A
5	44	60.3	40	14	US-10-029-386-32430, Sequence 32430, A
6	43	58.9	37	9	US-09-779-451-68, Sequence 58, Appl
7	43	58.9	37	12	US-09-803-060-43, Sequence 43, Appl
8	43	58.9	47	9	US-09-779-451-66, Sequence 66, Appl
9	43	58.9	47	12	US-09-803-060-41, Sequence 41, Appl
10	43	58.9	97	12	US-10-424-599-252376, Sequence 252376,
11	43	58.9	392	14	US-10-156-761-9175, Sequence 9175, Ap
12	43	58.9	938	14	US-10-174-677-100, Sequence 100, App
13	42	57.5	8	14	US-10-351-641-1664, Sequence 1664, Ap
14	42	57.5	27	14	US-10-106-638-5123, Sequence 5123, Ap
15	42	57.5	62	14	US-10-029-386-27875, Sequence 27875, A

16	42	57.5	143	12	US-10-282-122A-54902, Sequence 54902, A
17	42	57.5	143	15	US-10-289-762-6, Sequence 6, Appli
18	42	57.5	151	15	US-10-108-260A-3451, Sequence 3451, Ap
19	42	57.5	282	12	US-10-282-122A-63044, Sequence 63044, A
20	42	57.5	439	14	US-10-156-761-11997, Sequence 11997, A
21	42	57.5	930	16	US-10-408-765A-2639, Sequence 2639, Ap
22	41	56.2	20	15	US-10-72-437A-4, Sequence 4, Appli
23	41	56.2	53	12	US-10-424-599-217832, Sequence 217832,
24	41	56.2	109	9	US-09-989-920-223, Sequence 223, App
25	41	56.2	121	14	US-10-106-698-5667, Sequence 5667, Ap
26	41	56.2	124	12	US-10-425-114-43496, Sequence 43496, A
27	41	56.2	134	9	US-09-738-626-4920, Sequence 4920, Ap
28	41	56.2	210	14	US-10-002-631C-199, Sequence 199, App
29	41	56.2	223	12	US-10-424-599-164216, Sequence 164216,
30	41	56.2	340	14	US-10-107-857-16, Sequence 16, Appl
31	41	56.2	341	14	US-10-107-857-42, Sequence 42, Appl
32	41	56.2	342	15	US-10-074-978A-30, Sequence 30, Appl
33	41	56.2	372	14	US-10-107-857-41, Sequence 41, Appl
34	41	56.2	374	9	US-09-789-561-137, Sequence 137, App
35	41	56.2	374	11	US-09-833-245-2094, Sequence 2094, Ap
36	41	56.2	374	14	US-10-107-857-15, Sequence 15, Appl
37	41	56.2	374	15	US-10-333-900-25, Sequence 25, Appl
38	41	56.2	403	15	US-10-264-049-2819, Sequence 2819, Ap
39	41	56.2	493	15	US-10-369-493-5834, Sequence 5834, Ap
40	41	56.2	936	12	US-10-243-552-850, Sequence 850, App
41	40	54.8	30	9	US-09-925-299-1018, Sequence 1018, Ap
42	40	54.8	30	10	US-09-925-299-1028, Sequence 1028, Ap
43	40	54.8	60	9	US-09-867-550-1278, Sequence 1278, Ap
44	40	54.8	74	13	US-10-001-835-210, Sequence 210, App
45	40	54.8	75	13	US-10-001-835-182, Sequence 182, App
46	40	54.8	75	13	US-10-001-835-195, Sequence 195, App
47	40	54.8	75	13	US-10-001-835-223, Sequence 223, App
48	40	54.8	75	13	US-10-001-835-225, Sequence 225, App
49	40	54.8	105	14	US-10-029-386-32277, Sequence 32277, A
50	40	54.8	129	12	US-10-425-114-58367, A
51	40	54.8	130	15	US-10-198-260A-3005, Sequence 3005, Ap
52	40	54.8	176	14	US-10-106-698-5048, Sequence 5048, Ap
53	40	54.8	245	12	US-10-425-114-70663, A
54	40	54.8	303	12	US-10-282-122A-58144, Sequence 58144, A
55	40	54.8	318	12	US-10-282-122A-67471, Sequence 67471, A
56	40	54.8	329	12	US-10-425-114-53299, Sequence 53299, A
57	40	54.8	352	12	US-10-282-122A-59260, Sequence 59260, A
58	40	54.8	460	15	US-10-369-493-8708, Sequence 8708, Ap
59	40	54.8	591	15	US-10-369-493-16822, Sequence 16822, A
60	38	53.4	36	9	US-09-779-451-74, Sequence 74, Appl
61	39	53.4	36	12	US-09-809-060-49, Sequence 49, Appl
62	39	53.4	46	9	US-09-779-451-72, Sequence 72, Appl
63	39	53.4	46	12	US-10-106-698-4727, Sequence 4727, Ap
64	39	53.4	47	14	US-10-425-114-72688, Sequence 72688, A
65	39	53.4	55	12	US-10-425-114-72688, Sequence 72688, A
66	39	53.4	60	9	US-09-879-957-112, Sequence 112, App
67	39	53.4	64	12	US-10-424-599-161138, Sequence 161138,
68	39	53.4	71	14	US-10-163-587A-27, Sequence 27, Appl
69	39	53.4	73	11	US-09-864-408A-7046, Sequence 7046, Ap
70	39	53.4	108	14	US-10-263-103-33, Sequence 33, Appl
71	39	53.4	110	10	US-09-764-891-5219, Sequence 5219, Ap
72	39	53.4	119	12	US-10-424-599-171353, Sequence 171353,
73	39	53.4	127	11	US-09-833-245-577, Sequence 577, App
74	39	53.4	127	11	US-09-833-245-579, Sequence 579, App
75	39	53.4	134	12	US-10-424-599-281013, Sequence 281013,
76	39	53.4	144	9	US-09-879-957-6, Sequence 6, Appli
77	39	53.4	152	12	US-10-425-114-54667, Sequence 54667, A
78	39	53.4	157	14	US-10-080-170-116, Sequence 116, App
79	39	53.4	161	14	US-10-080-170-552, Sequence 552, App
80	39	53.4	206	12	US-10-424-599-218905, Sequence 218905,
81	39	53.4	243	9	US-09-854-816-73, Sequence 73, Appl
82	39	53.4	243	12	US-10-900-345A-108, Sequence 108, App
83	39	53.4	243	12	US-10-425-114-69876, Sequence 69876, A
84	39	53.4	243	14	US-10-305-765-142, Sequence 142, App
85	39	53.4	243	14	US-10-305-633-142, Sequence 142, Appl
86	39	53.4	282	9	US-09-247-890-16, Sequence 16, Appl
87	39	53.4	282	15	US-10-383-317-16, Sequence 16, Appl
88	39	53.4	351	14	US-10-026-741-47, Sequence 47, Appl

89	39	53.4	451	10	US-09-976-782-82	Sequence 82, Appl	162	38	52.1	40	14	US-10-351-641-1443	Sequence 1443, Ap
90	39	53.4	642	15	US-10-369-493-1005	Sequence 1005, Ap	163	38	52.1	40	14	US-10-351-641-1756	Sequence 1756, Ap
91	39	53.4	654	15	US-10-369-493-22744	Sequence 22744, A	164	38	52.1	45	14	US-10-086-409A-6	Sequence 6, Appli
92	39	53.4	715	14	US-10-320-786-134	Sequence 134, App	165	38	52.1	46	9	US-09-779-451-51	Sequence 51, Appl
93	39	53.4	766	15	US-10-094-749-2905	Sequence 2905, Ap	166	38	52.1	46	9	US-09-854-816-110	Sequence 110, App
94	39	53.4	842	14	US-10-220-674-2	Sequence 2, Appli	167	38	52.1	46	12	US-09-809-060-26	Sequence 26, App
95	39	53.4	842	15	US-10-297-022-1	Sequence 1, Appli	168	38	52.1	46	12	US-10-663-589-49	Sequence 49, Appl
96	39	53.4	842	15	US-10-295-027-1281	Sequence 1281, Ap	169	38	52.1	46	14	US-10-351-641-1089	Sequence 1089, Ap
97	39	53.4	858	14	US-10-190-435-150	Sequence 150, App	170	38	52.1	48	14	US-10-351-641-1090	Sequence 1090, Ap
98	39	53.4	862	14	US-10-369-294-15	Sequence 150, App	171	38	52.1	50	9	US-09-796-692-1205	Sequence 1205, Ap
99	39	53.4	877	14	US-10-026-741-102	Sequence 102, App	172	38	52.1	50	9	US-09-796-692-1205	Sequence 1205, Ap
100	39	53.4	912	9	US-09-817-464-2	Sequence 2, Appli	173	38	52.1	50	9	US-09-796-692-1205	Sequence 1205, Ap
101	39	53.4	912	14	US-10-225-567A-176	Sequence 176, App	174	38	52.1	50	14	US-10-040-862-1205	Sequence 1205, Ap
102	39	53.4	912	14	US-10-331-289-2	Sequence 2, Appli	175	38	52.1	50	14	US-10-040-862-1205	Sequence 1205, Ap
103	39	53.4	983	14	US-10-255-149-2	Sequence 2, Appli	176	38	52.1	50	14	US-10-040-862-1205	Sequence 1205, Ap
104	39	53.4	1002	15	US-10-262-445-66	Sequence 66, Appl	177	38	52.1	50	15	US-10-057-475B-1205	Sequence 1205, Ap
105	39	53.4	1005	13	US-10-114-091-2	Sequence 2, Appli	178	38	52.1	50	15	US-10-057-475B-1205	Sequence 1205, Ap
106	39	53.4	1005	15	US-10-262-445-62	Sequence 62, Appl	179	38	52.1	50	15	US-10-057-475B-1205	Sequence 1205, Ap
107	39	53.4	1068	15	US-10-262-445-64	Sequence 64, Appl	180	38	52.1	50	15	US-10-154-884B-1205	Sequence 1205, Ap
108	39	53.4	1446	12	US-10-343-649-7	Sequence 7, Appli	181	38	52.1	50	15	US-10-154-884B-1205	Sequence 1205, Ap
109	39	53.4	1593	12	US-10-343-649-6	Sequence 6, Appli	182	38	52.1	50	15	US-10-154-884B-1205	Sequence 1205, Ap
110	38.5	52.7	20	10	US-09-858-935B-86	Sequence 86, Appl	183	38	52.1	58	12	US-10-424-599-259464	Sequence 259464, Ap
111	38.5	52.7	20	12	US-10-271-869-86	Sequence 86, Appl	184	38	52.1	64	12	US-10-424-599-273932	Sequence 273932, Ap
112	38.5	52.7	34	9	US-09-864-761-45579	Sequence 45579, A	185	38	52.1	66	12	US-10-424-599-146752	Sequence 146752, Ap
113	38.5	52.7	794	9	US-09-738-626-4579	Sequence 4579, Ap	186	38	52.1	66	12	US-10-424-599-227257	Sequence 227257, Ap
114	38	52.1	8	14	US-10-351-641-1654	Sequence 1654, Ap	187	38	52.1	77	11	US-09-864-408A-5482	Sequence 5482, Ap
115	38	52.1	8	14	US-10-351-641-1662	Sequence 1662, Ap	188	38	52.1	80	14	US-10-106-698-5259	Sequence 5259, Ap
116	38	52.1	9	14	US-10-347-562-40	Sequence 24, Appl	189	38	52.1	108	12	US-10-424-599-221266	Sequence 221266, Ap
117	38	52.1	9	14	US-10-347-562-40	Sequence 40, Appl	190	38	52.1	132	12	US-10-425-114-36899	Sequence 36899, A
118	38	52.1	9	14	US-10-351-641-1564	Sequence 1564, Ap	191	38	52.1	147	12	US-10-424-599-167013	Sequence 167013, A
119	38	52.1	9	14	US-10-351-641-1580	Sequence 1580, Ap	192	38	52.1	160	12	US-10-425-114-70150	Sequence 70150, A
120	38	52.1	10	14	US-10-347-562-33	Sequence 33, Appl	193	38	52.1	172	12	US-10-424-599-238964	Sequence 238964, Ap
121	38	52.1	10	14	US-10-347-562-35	Sequence 35, Appl	194	38	52.1	191	12	US-10-424-599-263891	Sequence 263891, Ap
122	38	52.1	10	14	US-10-347-562-44	Sequence 44, Appl	195	38	52.1	210	12	US-10-425-114-53989	Sequence 53989, A
123	38	52.1	10	14	US-10-351-641-1573	Sequence 1573, Ap	196	38	52.1	221	14	US-10-156-761-9903	Sequence 9903, Ap
124	38	52.1	10	14	US-10-351-641-1575	Sequence 1575, Ap	197	38	52.1	242	12	US-10-424-599-239584	Sequence 239584, Ap
125	38	52.1	10	14	US-10-351-641-1584	Sequence 1584, Ap	198	38	52.1	268	9	US-09-854-816-68	Sequence 68, Appl
126	38	52.1	22	14	US-10-038-093-35	Sequence 35, Appl	199	38	52.1	268	9	US-09-854-816-70	Sequence 70, Appl
127	38	52.1	22	9	US-09-864-761-37150	Sequence 37150, A	200	38	52.1	268	9	US-09-854-816-72	Sequence 72, Appl
128	38	52.1	30	12	US-10-236-734-283	Sequence 283, App	201	38	52.1	269	9	US-09-854-816-25	Sequence 25, Appl
129	38	52.1	32	14	US-10-351-641-1193	Sequence 1193, Ap	202	38	52.1	269	9	US-09-854-816-33	Sequence 33, Appl
130	38	52.1	36	9	US-09-779-451-53	Sequence 53, Appl	203	38	52.1	270	9	US-09-854-816-67	Sequence 67, Appl
131	38	52.1	36	12	US-10-267-682-4	Sequence 4, Appli	204	38	52.1	320	12	US-10-424-599-206631	Sequence 206631, Ap
132	38	52.1	36	12	US-10-267-748-4	Sequence 4, Appli	205	38	52.1	432	14	US-10-424-599-200983	Sequence 200983, Ap
133	38	52.1	36	12	US-09-809-060-3	Sequence 3, Appli	206	38	52.1	432	14	US-10-101-464A-938	Sequence 938, App
134	38	52.1	36	12	US-08-809-060-28	Sequence 28, Appl	207	38	52.1	433	8	US-08-979-847-81	Sequence 81, Appl
135	38	52.1	36	12	US-10-663-589-39	Sequence 39, Appl	208	38	52.1	433	14	US-10-430-442-87	Sequence 87, Appl
136	38	52.1	36	14	US-10-252-136-4	Sequence 4, Appli	209	38	52.1	433	14	US-10-114-104-81	Sequence 81, Appl
137	38	52.1	36	14	US-10-351-641-1026	Sequence 1026, Ap	210	38	52.1	489	16	US-10-376-397A-4	Sequence 4, Appli
138	38	52.1	36	14	US-10-351-641-1027	Sequence 1027, Ap	211	38	52.1	501	15	US-10-369-493-6412	Sequence 6412, Ap
139	38	52.1	36	14	US-10-351-641-1028	Sequence 1028, Ap	212	38	52.1	504	12	US-10-282-122A-42934	Sequence 42934, A
140	38	52.1	36	14	US-10-351-641-1028	Sequence 1028, Ap	213	38	52.1	504	12	US-10-282-122A-69107	Sequence 69107, A
141	38	52.1	36	14	US-10-351-641-1353	Sequence 1353, Ap	214	38	52.1	505	12	US-10-282-122A-75583	Sequence 75583, A
142	38	52.1	36	14	US-10-351-641-1353	Sequence 1353, Ap	215	38	52.1	724	12	US-10-296-734-4485	Sequence 1485, Ap
143	38	52.1	36	14	US-10-351-641-1441	Sequence 1441, Ap	216	38	52.1	740	12	US-09-917-376-3	Sequence 3, Appli
144	38	52.1	37	14	US-10-351-641-1117	Sequence 1117, Ap	217	38	52.1	740	12	US-09-917-376-6	Sequence 6, Appli
145	38	52.1	38	14	US-10-086-409A-3	Sequence 3, Appli	218	38	52.1	740	14	US-10-155-400-3	Sequence 3, Appli
146	38	52.1	38	14	US-10-351-641-1166	Sequence 1166, Ap	219	38	52.1	740	14	US-10-155-400-6	Sequence 6, Appli
147	38	52.1	38	14	US-10-351-641-1167	Sequence 1167, Ap	220	38	52.1	860	14	US-10-190-435-6	Sequence 6, Appli
148	38	52.1	39	12	US-10-663-589-52	Sequence 52, Appl	221	38	52.1	860	14	US-10-190-435-146	Sequence 146, App
149	38	52.1	39	14	US-10-351-641-1098	Sequence 1098, Ap	222	38	52.1	860	14	US-10-241-009-6	Sequence 6, Appli
150	38	52.1	39	14	US-10-351-641-1131	Sequence 1131, Ap	223	38	52.1	860	14	US-10-190-434B-6	Sequence 6, Appli
151	38	52.1	39	14	US-10-351-641-1131	Sequence 1131, Ap	224	38	52.1	860	14	US-10-190-305A-6	Sequence 6, Appli
152	38	52.1	39	14	US-10-351-641-1150	Sequence 1150, Ap	225	38	52.1	862	14	US-10-190-435-141	Sequence 141, App
153	38	52.1	39	14	US-10-351-641-1151	Sequence 1151, Ap	226	38	52.1	865	14	US-10-190-435-140	Sequence 140, App
154	38	52.1	39	14	US-10-351-641-1152	Sequence 1152, Ap	227	38	52.1	867	14	US-10-190-435-3	Sequence 3, Appli
155	38	52.1	39	14	US-10-351-641-1165	Sequence 1165, Ap	228	38	52.1	867	14	US-10-190-435-126	Sequence 126, App
156	38	52.1	39	14	US-10-351-641-1270	Sequence 1270, Ap	229	38	52.1	867	14	US-10-241-009-3	Sequence 3, Appli
157	38	52.1	39	14	US-10-351-641-1273	Sequence 1273, Ap	230	38	52.1	867	14	US-10-190-434B-3	Sequence 3, Appli
158	38	52.1	39	14	US-10-351-641-1432	Sequence 1432, Ap	231	38	52.1	867	14	US-10-190-305A-3	Sequence 3, Appli
159	38	52.1	40	14	US-10-351-641-1438	Sequence 1438, Ap	232	38	52.1	869	14	US-10-190-435-4	Sequence 4, Appli
160	38	52.1	40	14	US-10-351-641-1439	Sequence 1439, Ap	233	38	52.1	869	14	US-10-241-009-4	Sequence 4, Appli
161	38	52.1	40	14	US-10-351-641-1442	Sequence 1442, Ap	234	38	52.1	869	14	US-10-190-434B-4	Sequence 4, Appli

235 38 52.1 869 14 US-10-190-305A-4 Sequence 4, Appli
236 38 52.1 870 14 US-10-190-435-127 Sequence 127, App
237 38 52.1 957 12 US-09-917-376-1 Sequence 1, Appli
238 38 52.1 957 12 US-10-155-400-1 Sequence 1, Appli
239 38 52.1 961 12 US-10-425-114-40733 Sequence 40733, A
240 38 52.1 963 12 US-10-425-114-62436 Sequence 62436, A
241 38 52.1 1277 10 US-09-934-455-428 Sequence 428, App
242 38 52.1 1661 15 US-10-320-797-3360 Sequence 3360, App
243 37.5 51.4 123 12 US-10-424-599-221252 Sequence 221252,
244 37.5 51.4 254 16 US-10-408-765A-1496 Sequence 1496, App
245 37.5 51.4 340 12 US-10-282-122A-51268 Sequence 51268, A
246 37.5 51.4 868 15 US-10-369-493-3985 Sequence 3985, App
247 37 50.7 8 14 US-10-351-641-1630 Sequence 1630, App
248 37 50.7 8 14 US-10-351-641-1645 Sequence 1645, App
249 37 50.7 8 14 US-10-351-641-1649 Sequence 1649, App
250 37 50.7 8 14 US-10-351-641-1657 Sequence 1657, App
251 37 50.7 9 14 US-10-347-562-42 Sequence 42, Appl
252 37 50.7 9 14 US-10-351-641-1582 Sequence 1582, App
253 37 50.7 10 14 US-10-347-562-38 Sequence 38, Appl
254 37 50.7 10 14 US-10-351-641-1578 Sequence 1578, App
255 37 50.7 11 9 US-09-030-619-56 Sequence 56, Appl
256 37 50.7 11 12 US-10-609-217-232 Sequence 232, App
257 37 50.7 11 12 US-10-609-217-251 Sequence 251, App
258 37 50.7 11 12 US-10-632-388-232 Sequence 232, App
259 37 50.7 11 12 US-10-632-388-251 Sequence 251, App
260 37 50.7 11 12 US-10-651-723-232 Sequence 232, App
261 37 50.7 11 12 US-10-651-723-251 Sequence 251, App
262 37 50.7 11 12 US-10-277-232-56 Sequence 56, Appl
263 37 50.7 11 12 US-10-645-761-232 Sequence 232, App
264 37 50.7 11 12 US-10-645-761-251 Sequence 251, App
265 37 50.7 11 14 US-10-229-368-52 Sequence 52, Appl
266 37 50.7 11 14 US-10-225-087-49 Sequence 49, Appl
267 37 50.7 11 15 US-10-277-233-56 Sequence 56, Appl
268 37 50.7 11 15 US-10-351-985-61 Sequence 61, Appl
269 37 50.7 11 16 US-10-666-696-232 Sequence 232, App
270 37 50.7 11 16 US-10-666-696-251 Sequence 251, App
271 37 50.7 11 16 US-10-653-048-232 Sequence 232, App
272 37 50.7 11 16 US-10-653-048-251 Sequence 251, App
273 37 50.7 13 12 US-10-267-682-172 Sequence 172, App
274 37 50.7 13 12 US-10-267-748-172 Sequence 172, App
275 37 50.7 13 14 US-10-252-136-25 Sequence 25, Appl
276 37 50.7 13 14 US-10-351-641-530 Sequence 530, App
277 37 50.7 19 10 US-09-999-220B-62 Sequence 62, Appl
278 37 50.7 20 10 US-09-999-220B-56 Sequence 56, Appl
279 37 50.7 36 10 US-09-999-220B-125 Sequence 125, App
280 37 50.7 38 14 US-10-351-641-1169 Sequence 1169, App
281 37 50.7 38 14 US-10-351-641-1170 Sequence 1170, App
282 37 50.7 39 12 US-10-663-589-59 Sequence 59, Appl
283 37 50.7 39 14 US-10-351-641-1168 Sequence 1168, App
284 37 50.7 39 14 US-10-351-641-1258 Sequence 1258, App
285 37 50.7 39 14 US-10-351-641-1261 Sequence 1261, App
286 37 50.7 39 14 US-10-351-641-1271 Sequence 1271, App
287 37 50.7 39 14 US-10-351-641-1272 Sequence 1272, App
288 37 50.7 45 9 US-09-864-761-39113 Sequence 39113, A
289 37 50.7 49 9 US-09-764-969-826 Sequence 826, App
290 37 50.7 49 14 US-10-091-504-826 Sequence 826, App
291 37 50.7 49 15 US-10-227-577-826 Sequence 826, App
292 37 50.7 57 12 US-10-424-599-172396 Sequence 172396,
293 37 50.7 57 14 US-10-029-386-28834 Sequence 28834, A
294 37 50.7 63 10 US-09-764-891-5347 Sequence 5347, App
295 37 50.7 64 15 US-10-264-049-2542 Sequence 2542, App
296 37 50.7 66 12 US-10-424-599-199225 Sequence 199225,
297 37 50.7 67 15 US-10-131-410-135 Sequence 135, App
298 37 50.7 85 12 US-10-425-114-37959 Sequence 37959, A
299 37 50.7 86 12 US-10-424-599-174560 Sequence 174560,
300 37 50.7 88 12 US-10-424-599-280347 Sequence 280347,

ALIGNMENTS

; Sequence 152002, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 152002
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(53)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_108281C.1.pep
US-10-424-599-152002

Query Match 72.6%; Score 53; DB 12; Length 53;
Best Local Similarity 85.7%; Pred. No. 7.4;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 WTNIWWW 8
| | | | |
Db 28 WINIWWW 34

RESULT 2
US-10-424-599-276989
; Sequence 276989, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 276989
; LENGTH: 88
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(88)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_9213C.1.pep
US-10-424-599-276989

Query Match 61.6%; Score 45; DB 12; Length 88;
Best Local Similarity 66.7%; Pred. No. 1e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 WTNIWWWAK 10
| | | | |
Db 43 WTVVWTWAK 51

RESULT 3
US-10-425-114-53192
; Sequence 53192, Application US/10425114

Publication No. US20040034888A1

GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 53192
LENGTH: 99
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: 700241044_FLI.pep
US-10-425-114-53192

Query Match 61.6%; Score 45; DB 12; Length 99;
Best Local Similarity 44.4%; Pred. No. 1.1e+02;
Matches 4; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 WTNWVWAK 10
| :|||:
Db 7 WCGTWWWSR 15

RESULT 4

US-10-425-114-47464
Sequence 47464, Application US/10425114
Publication No. US20040034888A1
GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 47464
LENGTH: 189
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: 700235331_FLI.pep
US-10-425-114-47464

Query Match 61.6%; Score 45; DB 12; Length 189;
Best Local Similarity 60.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 EWTNIWWAK 10
| :|||:
Db 176 EWRKIWLAR 185

RESULT 5

US-10-029-386-32430
Sequence 32430, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
FILE REFERENCE: AEMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 32430
LENGTH: 40
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC004065.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.5
US-10-029-386-32430

Query Match 60.3%; Score 44; DB 14; Length 40;
Best Local Similarity 57.1%; Pred. No. 72;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 WTNWVW 8
| :|||:
Db 15 WRWVWVW 21

RESULT 6

US-09-779-451-68
Sequence 68, Application US/09779451
Patent No. US20020094521A1
GENERAL INFORMATION:
APPLICANT: Wild, Carl T.
TITLE OF INVENTION: Assay for Detection of Viral Fusion Inhibitors
FILE REFERENCE: 1900.0300003
CURRENT APPLICATION NUMBER: US/09/779,451
CURRENT FILING DATE: 2001-08-17
PRIOR APPLICATION NUMBER: US 60/235,901
PRIOR FILING DATE: 2000-09-28
PRIOR APPLICATION NUMBER: US 60/181,543
PRIOR FILING DATE: 2000-02-10
NUMBER OF SEQ ID NOS: 77
SOFTWARE: PatentIn version 3.0
SEQ ID NO 68
LENGTH: 37
TYPE: PRT
ORGANISM: Human immunodeficiency virus type 1
US-09-779-451-68

Query Match 58.9%; Score 43; DB 9; Length 37;
Best Local Similarity 62.5%; Pred. No. 89;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNIWW 8
| :|||:
Db 28 KWTNLANW 35

RESULT 7

US-09-809-060-43
Sequence 43, Application US/09809060
Publication No. US20020010317A1
GENERAL INFORMATION:
APPLICANT: Wild, Carl T.
TITLE OF INVENTION: A Method for Generating Immunogens that Elicit
Neutralizing Antibodies against Fusion-Active
TITLE OF INVENTION: Regions of HIV Envelope Proteins
FILE REFERENCE: 1900.0280001
CURRENT APPLICATION NUMBER: US/09/809,060
CURRENT FILING DATE: 2001-03-16

; PRIOR APPLICATION NUMBER: US 60/189,981
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 43
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-809-060-43

Query Match 58.9%; Score 43; DB 12; Length 37;
Best Local Similarity 62.5%; Pred. No. 89;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNIWW 8
Db 28 KWTNLWN 35

RESULT 8
US-09-779-451-66
; Sequence 66, Application US/09779451
; Patent No. US2002009452A1
; GENERAL INFORMATION:
; APPLICANT: Wild, Carl T.
; TITLE OF INVENTION: Assay for Detection of Viral Fusion Inhibitors
; FILE REFERENCE: 1900.030003
; CURRENT APPLICATION NUMBER: US/09/779,451
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 60/235,901
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: JS 60/181,543
; PRIOR FILING DATE: 2000-02-13
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 66
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-779-451-66

Query Match 58.9%; Score 43; DB 9; Length 47;
Best Local Similarity 62.5%; Pred. No. 11e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNIWW 8
Db 38 KWTNLWN 45

RESULT 9
US-09-809-060-41
; Sequence 41, Application US/09809060
; Publication No. US20020010317A1
; GENERAL INFORMATION:
; APPLICANT: Wild, Carl T.
; TITLE OF INVENTION: A Method for Generating Immunogens that Elicit
; TITLE OF INVENTION: Neutralizing Antibodies against Fusion-Active
; TITLE OF INVENTION: Regions of HIV Envelope Proteins
; FILE REFERENCE: 1900.0260001
; CURRENT APPLICATION NUMBER: US/09/809,060
; CURRENT FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: US 60/189,981
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 41
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-809-060-41

Query Match 58.9%; Score 43; DB 12; Length 47;
Best Local Similarity 62.5%; Pred. No. 11e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNIWW 8
Db 38 KWTNLWN 45

RESULT 10
US-10-424-599-252376
; Sequence 252376, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 252376
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_69924C.1.pap
US-10-424-599-252376

Query Match 58.9%; Score 43; DB 12; Length 97;
Best Local Similarity 50.6%; Pred. No. 1.9e+02;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 EWTNIWWAK 10
Db 26 KWTTFWIK 35

RESULT 11
US-10-156-761-9175
; Sequence 9175, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9175
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9175

Query Match 58.9%; Score 43; DB 14; Length 392;
Best Local Similarity 83.3%; Pred. No. 5.7e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TWNIWWW 8
|||
Db 324 TNAWWW 329

RESULT 12

US-10-174-677-100
; Sequence 100, Application US/10174677
; Publication No. US20030190704A1
; GENERAL INFORMATION:
; APPLICANT: Xie, Ting
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ANCHORING STEM CELLS IN A MICROENVIR
; FILE REFERENCE: 40716(IP-012)
; CURRENT APPLICATION NUMBER: US/10/174,677
; CURRENT FILING DATE: 2002-06-19
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 100
; LENGTH: 938
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-174-677-100

Query Match 58.9%; Score 43; DB 14; Length 938;

Best Local Similarity 75.0%; Pred. No. 1.1e+03;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 WTNLWWWA 9
|||
Db 8 WTELRWA 15

RESULT 13

US-10-351-641-1664
; Sequence 1664, Application US/10351641
; Publication No. US20030186874A1
; GENERAL INFORMATION:
; APPLICANT: Barney, S.
; APPLICANT: Guthrie, K.
; APPLICANT: Merutka, G.
; APPLICANT: Anwer, M.
; APPLICANT: Lambert, D.

; TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED PHARMACOKINETIC
; TITLE OF INVENTION: PROPERTIES

; FILE REFERENCE: 7872-100
; CURRENT APPLICATION NUMBER: US/10/351,641
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: 09/350,641
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/315,304
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: 09/082,279
; PRIOR FILING DATE: 1998-05-20
; NUMBER OF SEQ ID NOS: 1757
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1664
; LENGTH: 8
; TYPE: PRT
; ORGANISM: HIV-1
US-10-351-641-1664

Query Match 57.5%; Score 42; DB 14; Length 8;

Best Local Similarity 71.4%; Pred. No. 1e+06;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 WTNLWWWA 8
|||
Db 1 WTNLWWN 7

RESULT 14

US-10-106-698-5123
; Sequence 5123, Application US/10106698

; Publication No. US20030109690A1

; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypepti
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5123
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (25)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-5123

Query Match 57.5%; Score 42; DB 14; Length 27;

Best Local Similarity 75.0%; Pred. No. 92;
Matches 6; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 2 WTNLWWWA 9
|||
Db 12 WT--WWWA 17

RESULT 15

US-10-029-386-27875
; Sequence 27875, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:

; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David K.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AECMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 27875
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL137918.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.94
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.62
; OTHER INFORMATION: SWISSPROT HIT: O55029, EVALU8 3.50e+00
US-10-029-386-27875

Query Match 57.5%; Score 42; DB 14; Length 62;

Best Local Similarity 55.6%; Pred. No. 1.8e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNWWWA 9
|:|
Db 27 ETSRWWS 35

```
RESULT 16
US-10-282-122A-54902
; Sequence 54902, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Chlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54902
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-10-282-122A-54902

Query Match          57.5%; Score 42; DB 12; Length 143;
Best Local Similarity 66.7%; Pred. No. 3.4e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 WTNWVW 7
Db      117 WSNLWV 122

RESULT 17
US-10-289-762-6
; Sequence 6, Application US/10289762
; Publication No. US20040006218A1
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/10/289,762
; PRIOR FILING DATE: 2003-03-27
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 6
```

```
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-10-289-762-6

Query Match          57.5%; Score 42; DB 15; Length 143;
Best Local Similarity 66.7%; Pred. No. 3.4e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      2 WTNWVW 7
Db      117 WSNLWV 122

RESULT 18
US-10-108-260A-3451
; Sequence 3451, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 3451
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3451

Query Match          57.5%; Score 42; DB 15; Length 151;
Best Local Similarity 57.1%; Pred. No. 3.6e+02;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      4 NIWVWAK 10
Db      106 SVWVWVWK 112

RESULT 19
US-10-282-122A-63044
; Sequence 63044, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
```

; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 63044
; LENGTH: 282
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-10-282-122A-63044

Query Match 57.5%; Score 42; DB 12; Length 282;
Best Local Similarity 62.5%; Pred. No. 5.8e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WTNIWWA 9
DB 72 WTWVFWA 79

RESULT 20
US-10-156-761-11997
; Sequence 11997, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: CMJRA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204889
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272897
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11997
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11997

Query Match 57.5%; Score 42; DB 14; Length 439;
Best Local Similarity 57.1%; Pred. No. 8.3e+02;
Matches 4; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 WTNIWWA 8
DB 290 WARSWWA 296

RESULT 21
US-10-408-765A-2639
; Sequence 2639, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.

; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2639
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2639

Query Match 57.5%; Score 42; DB 16; Length 930;
Best Local Similarity 54.5%; Pred. No. 1.5e+03;
Matches 6; Conservative 3; Mismatches 0; Indels 2; Gaps 0;

QY 2 WTNI--WWWAX 10
DB 24 WTNLHKWPFK 34

RESULT 22
US-10-272-437A-4
; Sequence 4, Application US/10272437A
; Publication No. US20030216309A1
; GENERAL INFORMATION:
; APPLICANT: Krag, David N.
; APPLICANT: Oligino, Lyn
; TITLE OF INVENTION: BINDING PEPTIDES SPECIFIC FOR THE EXTRACELLULAR DOMAIN OF ERBB2
; FILE REFERENCE: V00139.70056.US
; CURRENT APPLICATION NUMBER: US/10/272,437A
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: 60/329,183
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 4
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide
US-10-272-437A-4

Query Match 56.2%; Score 41; DB 15; Length 20;
Best Local Similarity 62.5%; Pred. No. 96;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNIWWW 8
DB 8 EWQNDWFW 15

RESULT 23
US-10-424-599-217832
; Sequence 217832, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: Ia Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684

```
; SEQ ID NO 217832
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_38729C.1.pep
US-10-424-599-217832

Query Match
Best Local Similarity 56.2%; Score 41; DB 12; Length 53;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 WTNLWVAK 10
Db 24 WTRIWTWR 32

RESULT 24
US-09-989-920-223
; Sequence 223, Application US/09989920
; Patent No. US20020172957A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Hervé
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Proteins
; FILE REFERENCE: DEX-0291
; CURRENT APPLICATION NUMBER: US/09/989,920
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/252,500
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 284
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 223
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-989-920-223

Query Match
Best Local Similarity 56.2%; Score 41; DB 9; Length 109;
Matches 5; Conservative 2; Mismatches 1; Indels 4; Gaps 1;

Qy 1 EWTN-----IWMW 8
Db 12 EWSEMLGVWVWVW 23

RESULT 25
US-10-106-698-5667
; Sequence 5667, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA0035P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5667
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-5667

; SEQ ID NO 217832
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_38729C.1.pep
US-10-424-599-217832

Query Match
Best Local Similarity 56.2%; Score 41; DB 12; Length 53;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 WTNLWVAK 10
Db 24 WTRIWTWR 32

RESULT 24
US-09-989-920-223
; Sequence 223, Application US/09989920
; Patent No. US20020172957A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Hervé
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Proteins
; FILE REFERENCE: DEX-0291
; CURRENT APPLICATION NUMBER: US/09/989,920
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/252,500
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 284
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 223
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-989-920-223

Query Match
Best Local Similarity 56.2%; Score 41; DB 9; Length 109;
Matches 5; Conservative 2; Mismatches 1; Indels 4; Gaps 1;

Qy 1 EWTN-----IWMW 8
Db 12 EWSEMLGVWVWVW 23

RESULT 25
US-10-106-698-5667
; Sequence 5667, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA0035P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5667
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-5667

; SEQ ID NO 217832
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_38729C.1.pep
US-10-424-599-217832

Query Match
Best Local Similarity 56.2%; Score 41; DB 14; Length 121;
Matches 4; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 WTNLWVW 8
Db 17 WQGRVWVW 23

RESULT 26
US-10-425-114-43496
; Sequence 43496, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 43496
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700349016_FLI.pep
US-10-425-114-43496

Query Match
Best Local Similarity 56.2%; Score 41; DB 12; Length 124;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 IWWVA 9
Db 31 IWWVA 35

RESULT 27
US-09-738-626-4920
; Sequence 4920, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4920
; LENGTH: 134
```

```
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4920

Query Match      56.2%; Score 41; DB 9; Length 134;
Best Local Similarity 56.7%; Pred. No. 4.3e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 5 LWWAX 10
Db 96 LWWAR 101

RESULT 28
US-10-002-631C-199
; Sequence 199, Application US/10002631C
; Publication No. US20030157486A1
; GENERAL INFORMATION:
; APPLICANT: Graff, Jonathan M.
; APPLICANT: Muenster, Matthew
; TITLE OF INVENTION: METHODS TO IDENTIFY SIGNAL SEQUENCES
; FILE REFERENCE: A34943 090495.0243
; CURRENT APPLICATION NUMBER: US/10/002.631C
; CURRENT FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/300,309
; PRIOR FILING DATE: 2001-06-21
; NUMBER OF SEQ ID NOS: 324
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 193
; LENGTH: 213
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (6)...(6)
; OTHER INFORMATION: Xaa = any amino acid
US-10-002-631C-199

Query Match      56.2%; Score 41; DB 14; Length 210;
Best Local Similarity 71.4%; Pred. No. 6.1e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNIW 7
Db 177 EWTSDW 183

RESULT 29
US-10-424-599-164216
; Sequence 164216, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)/B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 164216
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(223)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_119305C.1.pap
US-10-424-599-164216
```

```
Query Match      56.2%; Score 41; DB 12; Length 223;
Best Local Similarity 71.4%; Pred. No. 6.4e+02;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TNIWWA 9
Db 49 TTFWWA 55

RESULT 30
US-10-107-857-16
; Sequence 16, Application US/10107857
; Publication No. US20030170784A1
; GENERAL INFORMATION:
; APPLICANT: Fraser, Christopher C.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 07334-251001
; CURRENT APPLICATION NUMBER: US/10/107,857
; CURRENT FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: 09/514,009
; PRIOR FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-107-857-16

Query Match      56.2%; Score 41; DB 14; Length 340;
Best Local Similarity 71.4%; Pred. No. 8.9e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNIW 7
Db 266 EWTSDW 272

RESULT 31
US-10-107-857-42
; Sequence 42, Application US/10107857
; Publication No. US20030170784A1
; GENERAL INFORMATION:
; APPLICANT: Fraser, Christopher C.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 07334-251001
; CURRENT APPLICATION NUMBER: US/10/107,857
; CURRENT FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: 09/514,009
; PRIOR FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Mus musculus
; US-10-107-857-42

Query Match      56.2%; Score 41; DB 14; Length 341;
Best Local Similarity 71.4%; Pred. No. 8.9e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 EWTNIW 7
Db 267 EWTSDW 273

RESULT 32
US-10-074-978A-30
; Sequence 30, Application US/10074978A
; Publication No. US20040010119A1
; GENERAL INFORMATION:
```


APPLICANT: Leite, Mario
APPLICANT: Spytek, Kimberly A
APPLICANT: Guo, Xiaojia (Sasha)
APPLICANT: Fernandes, Elma
APPLICANT: Li, Li
APPLICANT: Kekuda, Ramesh
APPLICANT: Liu, Xiaohong
APPLICANT: Casman, Stacie
APPLICANT: Boldog, Ferenc
APPLICANT: Patturajan, Meera
APPLICANT: Bialock, Angela
APPLICANT: Ballinger, Robert
APPLICANT: Vernet, Corine
APPLICANT: Tchernev, Velizar T
APPLICANT: Malyankar, Uriel M
APPLICANT: Gusev, Vladimir
APPLICANT: Rastelli, Luca
APPLICANT: Mezes, Peter S
APPLICANT: Ellerman, Karen
APPLICANT: Hayes, Melvin P
APPLICANT: Herrman, John
APPLICANT: Pena, Carol E A
APPLICANT: Shimkets, Richard A
APPLICANT: Taupier Jr, Raymond J
APPLICANT: Moore, No. US20040010119A11le
APPLICANT: Shenoy, Suresh
APPLICANT: Edinger, Shlomit
APPLICANT: Gunther, Erik
APPLICANT: Stone, Dave
APPLICANT: Millet, Isabelle
APPLICANT: Peyman, John
APPLICANT: Smithson, Glenda
TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: 21402-269
CURRENT APPLICATION NUMBER: US/10/074,978A
CURRENT FILING DATE: 2003-01-07
PRIOR APPLICATION NUMBER: 60/268,221
PRIOR FILING DATE: 2001-02-12
PRIOR APPLICATION NUMBER: 60/335,109
PRIOR FILING DATE: 2001-10-31
PRIOR APPLICATION NUMBER: 60/312,284
PRIOR FILING DATE: 2001-08-14
PRIOR APPLICATION NUMBER: 60/268,496
PRIOR FILING DATE: 2001-02-13
PRIOR APPLICATION NUMBER: 60/276,703
PRIOR FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: 60/330,293
PRIOR FILING DATE: 2001-10-18
PRIOR APPLICATION NUMBER: 60/322,127
PRIOR FILING DATE: 2001-11-21
PRIOR APPLICATION NUMBER: 60/280,899
PRIOR FILING DATE: 2001-04-02
PRIOR APPLICATION NUMBER: 60/310,797
PRIOR FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: 60/268,646
PRIOR FILING DATE: 2001-02-14
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 547
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 30
LENGTH: 344
TYPE: PRT
ORGANISM: Homo sapiens
US-10-074-978A-30

Query Match 56.2%; Score 41; DB 15; Length 344;
Best Local Similarity 57.1%; Pred. No. 9e+02;
Matches 4; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 WNTNWW 8
| | | |
Db 274 WDTSNWW 280

RESULT 33
US-10-107-857-41
; Sequence 41, Application US/10107857
; Publication No. US20030170784A1
; GENERAL INFORMATION:
; APPLICANT: Fraser, Christopher C.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 07334-251001
; CURRENT APPLICATION NUMBER: US/10/107,857
; CURRENT FILING DATE: 2002-03-26
; PRIOR APPLICATION NUMBER: 09/514,009
; PRIOR FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)...(31)
US-10-107-857-41

Query Match 56.2%; Score 41; DB 14; Length 372;
Best Local Similarity 71.4%; Pred. No. 9.6e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EWTNIWW 7
| | | |
Db 298 EWTSDWW 304

RESULT 34
US-09-789-561-137
; Sequence 137, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043P1
; CURRENT APPLICATION NUMBER: US/09/789,561
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 137
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-789-561-137

Query Match 56.2%; Score 41; DB 9; Length 374;
Best Local Similarity 71.4%; Pred. No. 9.6e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 EWTNIWW 7
| | | |
Db 300 EWTSDWW 306

RESULT 35
US-09-833-245-2094
; Sequence 2094, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins

```
? FILE REFERENCE: PF546PCT
? CURRENT APPLICATION NUMBER: US/09/833,245
? CURRENT FILING DATE: 2001-04-12
? PRIOR APPLICATION NUMBER: 60/229, 358
? PRIOR FILING DATE: 2000-04-12
? PRIOR APPLICATION NUMBER: 60/256, 931
? PRIOR FILING DATE: 2000-12-21
? PRIOR APPLICATION NUMBER: 60/199, 384
? PRIOR FILING DATE: 2000-04-25
? NUMBER OF SEQ ID NOS: 2267
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 2094
? LENGTH: 374
? TYPE: PRT
? ORGANISM: Homo sapiens
US-09-833-245-2094

Query Match      56.2%; Score 41; DB 11; Length 374;
Best Local Similarity 71.4%; Pred. No. 9.6e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 EWTNIWM 7
Db      300 EWTSDWM 306

RESULT 36
US-10-107-857-15
? Sequence 15, Application US/10107857
? Publication No. US20030170784A1
? GENERAL INFORMATION:
? APPLICANT: Fraser, Christopher C.
? TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
? FILE REFERENCE: 07334-251001
? CURRENT APPLICATION NUMBER: US/10/107,857
? CURRENT FILING DATE: 2002-03-26
? PRIOR APPLICATION NUMBER: 09/514,009
? PRIOR FILING DATE: 2000-02-25
? NUMBER OF SEQ ID NOS: 51
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 15
? LENGTH: 374
? TYPE: PRT
? ORGANISM: Homo sapiens
US-10-107-857-15

Query Match      56.2%; Score 41; DB 14; Length 374;
Best Local Similarity 71.4%; Pred. No. 9.6e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 EWTNIWM 7
Db      300 EWTSDWM 306

RESULT 37
US-10-333-900-25
? Sequence 25, Application US/10333900
? Publication No. US20040010121A1
? GENERAL INFORMATION:
? APPLICANT: Human Genome Sciences, Inc.
? TITLE OF INVENTION: 7 Human Ovarian and Ovarian Cancer Associated Proteins
? FILE REFERENCE: PA006PCT
? CURRENT APPLICATION NUMBER: US/10/333,900
? CURRENT FILING DATE: 2001-03-15
? PRIOR APPLICATION NUMBER: 60/227,009
? PRIOR FILING DATE: 2000-08-23
? PRIOR APPLICATION NUMBER: 60/190,076
? PRIOR FILING DATE: 2000-03-17
? NUMBER OF SEQ ID NOS: 34
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 25
? LENGTH: 374
```

```
? TYPE: PRT
? ORGANISM: Homo sapiens
US-10-333-900-25

Query Match      56.2%; Score 41; DB 15; Length 374;
Best Local Similarity 71.4%; Pred. No. 9.6e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 EWTNIWM 7
Db      300 EWTSDWM 306

RESULT 38
US-10-264-049-2819
? Sequence 2819, Application US/10264049
? Publication No. US20040005579A1
? GENERAL INFORMATION:
? APPLICANT: Birse et al.
? TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
? FILE REFERENCE: PA133P1
? CURRENT APPLICATION NUMBER: US/10/264,049
? CURRENT FILING DATE: 2002-10-04
? PRIOR APPLICATION NUMBER: PCT/US01/18569
? PRIOR FILING DATE: 2001-06-07
? PRIOR APPLICATION NUMBER: US 60/209,467
? PRIOR FILING DATE: 2000-06-07
? NUMBER OF SEQ ID NOS: 4360
? SOFTWARE: PatentIn Ver. 3.1
? SEQ ID NO 2819
? LENGTH: 403
? TYPE: PRT
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: MISC_FEATURE
? LOCATION: (4)
? OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
? NAME/KEY: MISC_FEATURE
? LOCATION: (5)
? OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
? FEATURE:
? NAME/KEY: MISC_FEATURE
? LOCATION: (370)
? OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-2819

Query Match      56.2%; Score 41; DB 15; Length 403;
Best Local Similarity 71.4%; Pred. No. 1e+03;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 EWTNIWM 7
Db      329 EWTSDWM 335

RESULT 39
US-10-369-493-5834
? Sequence 5834, Application US/10369493
? Publication No. US20030233675A1
? GENERAL INFORMATION:
? APPLICANT: Cao, Yongwei
? APPLICANT: Hinkle, Gregory J.
? APPLICANT: Slater, Steven C.
? APPLICANT: Goldman, Barry S.
? APPLICANT: Chen, Xianfeng
? TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
? FILE REFERENCE: 38-10(52052)B
? CURRENT APPLICATION NUMBER: US/10/369,493
? CURRENT FILING DATE: 2003-02-28
? PRIOR APPLICATION NUMBER: US 60/360,039
? PRIOR FILING DATE: 2002-02-21
```

```
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5834
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5834

Query Match          56.2%; Score 41; DB 15; Length 493;
Best Local Similarity 57.1%; Pred. No. 1.2e+03;
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      2 WTNINWV 8
Db      218 WTAVMWY 224

RESULT 40
US-10-243-552-850
; Sequence 850, Application US/10243552
; Publication No. US20030224379A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Weng, Gezhi
; APPLICANT: Ma, Yunqing
; TITLE OF INVENTION: Novel Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 807A
; CURRENT APPLICATION NUMBER: US/10/243.552
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/322,511
; PRIOR FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/04927
; PRIOR FILING DATE: 2001-02-26
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 998
; SOFTWARE: Pct_Fl_genes Version 5.0
; SEQ ID NO 850
; LENGTH: 936
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-243-552-850

Query Match          56.2%; Score 41; DB 12; Length 936;
Best Local Similarity 71.4%; Pred. No. 2e+03;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 EWTNINW 7
Db      845 EWTSDWV 851

RESULT 41
US-09-925-299-1018
; Sequence 1018, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: P102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1018
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-299-1018

Query Match          54.8%; Score 40; DB 9; Length 30;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      5 IWWWA 9
Db      13 VWWWA 17

RESULT 42
US-09-925-299-1018
; Sequence 1018, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: P102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1018
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-299-1018

Query Match          54.8%; Score 40; DB 10; Length 30;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      5 IWWWA 9
Db      13 VWWWA 17

RESULT 43
US-09-867-550-1278
; Sequence 1278, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1el Polynucleotides from Atherogenic Cells and
; FILE REFERENCE: 21402-013 [Cura-313]
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
```

; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 1278
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (35)
; OTHER INFORMATION: Wherein Xaa may be any one of Leu or Met or Val
; NAME/KEY: VARIANT
; LOCATION: (55)
; OTHER INFORMATION: Wherein Xaa may be any one of Asn or Asp or His or Tyr
US-09-867-550-1278

Query Match 54.8%; Score 40; DB 9; Length 60;
Best Local Similarity 33.3%; Pred. No. 3e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 6; Gaps 1;

QY 1 EWTNI-----WWNA 9
|| : |||:
Db 9 EWAQVLENYVQWWS 23

RESULT 44
US-10-001-835-210
; Sequence 210, Application US/10001835
; Publication No. US20020160387A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Caffrkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Pro
; FILE REFERENCE: DEX-0277
; CURRENT APPLICATION NUMBER: US/10/001,835
; PRIOR FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,997
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 210
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-835-210

Query Match 54.8%; Score 40; DB 13; Length 74;
Best Local Similarity 30.0%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 5 INWNA 9
: ||| |
Db 48 VWWNA 52

RESULT 45
US-10-001-835-182
; Sequence 182, Application US/10001835
; Publication No. US20020160387A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Caffrkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Pro
; FILE REFERENCE: DEX-0277

; CURRENT APPLICATION NUMBER: US/10/001,835
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,997
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 182
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-835-182

Query Match 54.8%; Score 40; DB 13; Length 75;
Best Local Similarity 80.0%; Pred. No. 3.6e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 5 INWNA 9
: ||| |
Db 48 VWWNA 52

RESULT 46
US-10-001-835-195
; Sequence 195, Application US/10001835
; Publication No. US20020160387A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Caffrkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and F
; FILE REFERENCE: DEX-0277
; CURRENT APPLICATION NUMBER: US/10/001,835
; PRIOR FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,997
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 195
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-835-195

Query Match 54.8%; Score 40; DB 13; Length 75;
Best Local Similarity 80.0%; Pred. No. 3.6e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 5 INWNA 9
: ||| |
Db 48 VWWNA 52

RESULT 47
US-10-001-835-223
; Sequence 223, Application US/10001835
; Publication No. US20020160387A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Caffrkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and P
; FILE REFERENCE: DEX-0277
; CURRENT APPLICATION NUMBER: US/10/001,835
; PRIOR FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,997
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 228

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 223
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-835-223

Query Match          54.8%; Score 40; DB 13; Length 75;
Best Local Similarity 80.0%; Pred. No. 3.6e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      5 IWWWA 9
        :||||
Db       48 VWWWA 52

RESULT 48
US-10-001-835-225
; Sequence 225, Application US/10001835
; Publication No. US20020160387A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Caffierkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Pro
; FILE REFERENCE: DEX-0277
; CURRENT APPLICATION NUMBER: US/10/001,835
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,997
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 225
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-835-225

Query Match          54.8%; Score 40; DB 13; Length 75;
Best Local Similarity 80.0%; Pred. No. 3.6e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      5 IWWWA 9
        :||||
Db       48 VWWWA 52

RESULT 49
US-10-029-386-32277
; Sequence 32277, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AECOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32277
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: KAP TO AC009336.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 223
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-835-223

Query Match          54.8%; Score 40; DB 13; Length 75;
Best Local Similarity 80.0%; Pred. No. 3.6e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      5 IWWWA 9
        :||||
Db       48 VWWWA 52

RESULT 48
US-10-001-835-225
; Sequence 225, Application US/10001835
; Publication No. US20020160387A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Caffierkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Pro
; FILE REFERENCE: DEX-0277
; CURRENT APPLICATION NUMBER: US/10/001,835
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,997
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 225
; LENGTH: 75
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-835-225

Query Match          54.8%; Score 40; DB 13; Length 75;
Best Local Similarity 80.0%; Pred. No. 3.6e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      5 IWWWA 9
        :||||
Db       48 VWWWA 52

RESULT 49
US-10-029-386-32277
; Sequence 32277, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AECOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32277
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: KAP TO AC009336.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
```

```
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.9
US-10-029-386-32277
```

```
Query Match          54.8%; Score 40; DB 14; Length 105;
Best Local Similarity 57.1%; Pred. No. 4.7e+02;
Matches 4; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      2 WTNIWWW 8
        :||||
Db       63 WRRRWWW 69
```

RESULT 50

```
US-10-425-114-58367
; Sequence 58367, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 58367
; LENGTH: 129
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3591-091-B12_FLI pep
US-10-425-114-58367
```

```
Query Match          54.8%; Score 40; DB 12; Length 129;
Best Local Similarity 57.1%; Pred. No. 5.5e+02;
Matches 4; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      2 WTNIWWW 8
        :||||
Db       21 WRRRWWW 27
```

RESULT 51

```
US-10-108-260A-3005
; Sequence 3005, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
; FILE REFERENCE: HL-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3005
; LENGTH: 130
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3005
```

```
Query Match          54.8%; Score 40; DB 15; Length 130;
Best Local Similarity 71.4%; Pred. No. 5.5e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      2 WTNIWWW 8
        :||||
```

```
Db      26 WTLIFWW 32

RESULT 52
US-10-106-698-5048
; Sequence 5048, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA05P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 52/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: Patent In Ver. 3.0
; SEQ ID NO 5048
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-5048

Query Match      54.8%; Score 40; DB 14; Length 176;
Best Local Similarity 80.0%; Pred. No. 7e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      5 IWWNA 9
      :|||
Db      22 VWWNA 26

RESULT 53
US-10-425-114-70663
; Sequence 70663, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhoul, Yihua
; APPLICANT: Kovalic, David X.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 70663
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73064A07_F11.pep
US-10-425-114-70663

Query Match      54.8%; Score 40; DB 12; Length 245;
Best Local Similarity 57.1%; Pred. No. 9.1e+02;
Matches 4; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 WTNWWWW 8
      :|||
Db      21 WRRWWWW 27

RESULT 54
US-10-282-122A-58144
; Sequence 58144, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
```

```
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: SLITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 58144
; LENGTH: 303
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-282-122A-58144

Query Match      54.8%; Score 40; DB 12; Length 303;
Best Local Similarity 66.7%; Pred. No. 1.1e+03;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      3 TNNWWW 8
      :|||
Db      124 TGLWWW 129

RESULT 55
US-10-282-122A-67471
; Sequence 67471, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
```

1 TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

2 FILE REFERENCE: ELITRA.034A

3 CURRENT APPLICATION NUMBER: US/10/282,122A

4 PRIOR FILING DATE: 2003-02-20

5 PRIOR FILING DATE: 2000-03-21

6 PRIOR FILING DATE: 2000-05-23

7 PRIOR FILING DATE: 2000-05-26

8 PRIOR FILING DATE: 2000-09-06

9 PRIOR FILING DATE: 2000-09-29

10 PRIOR FILING DATE: 2000-10-23

11 PRIOR FILING DATE: 2000-11-27

12 PRIOR FILING DATE: 2000-12-22

13 PRIOR FILING DATE: 2001-02-09

14 PRIOR FILING DATE: 2001-02-16

15 Remaining Prior Application data removed - See File Wrapper or PALM.

16 NUMBER OF SEQ ID NOS: 78614

17 SOFTWARE: Patent in version 3.1

18 SEQ ID NO 67471

19 LENGTH: 318

20 TYPE: PRT

21 ORGANISM: Pasteurella multocida

22 US-10-282-122A-67471

23 Query Match 54.8%; Score 40; DB 12; Length 318;

24 Best Local Similarity 66.7%; Pred. No. 1.1e+03;

25 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

26 QY 3 TNIWWW 8

27 DB 128 TGLWWW 133

28 RESULT 56

29 US-10-425-114-53299

30 Sequence 53299, Application US/10425114

31 Publication No. US20040034888A1

32 GENERAL INFORMATION:

33 APPLICANT: Liu, Jingdong

34 APPLICANT: Zhou, Yihua

35 APPLICANT: Kovalic, David K.

36 APPLICANT: Screen, Steven E.

37 APPLICANT: Tabaska, Jack E.

38 APPLICANT: Cao, Yongwei

39 TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

40 FILE REFERENCE: 38-21(53313)B

41 CURRENT APPLICATION NUMBER: US/10/425,114

42 CURRENT FILING DATE: 2003-04-28

43 NUMBER OF SEQ ID NOS: 73128

44 SEQ ID NO 53299

45 LENGTH: 329

46 TYPE: PRT

47 ORGANISM: Zea mays

48 FEATURE:

49 OTHER INFORMATION: Clone ID: 700240195_FLI.pap

50 US-10-425-114-53299

51 Query Match 54.8%; Score 40; DB 12; Length 329;

52 Best Local Similarity 50.0%; Pred. No. 1.1e+03;

53 Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

54 QY 1 EWTNIWWW 8

55 DB 128 TGLWWW 133

56 RESULT 57

57 US-10-282-122A-59260

58 Sequence 59260, Application US/10282122A

59 Publication No. US20040029129A1

60 GENERAL INFORMATION:

61 APPLICANT: Wang, Liangsu

62 APPLICANT: Zamudio, Carlos

63 APPLICANT: Malone, Cheryl

64 APPLICANT: Haselbeck, Robert

65 APPLICANT: Ohlsen, Kari

66 APPLICANT: Zyskind, Judith

67 APPLICANT: Wall, Daniel

68 APPLICANT: Trawick, John

69 APPLICANT: Carr, Grant

70 APPLICANT: Yamamoto, Robert

71 APPLICANT: Porsyth, R.

72 TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

73 FILE REFERENCE: ELITRA.034A

74 CURRENT APPLICATION NUMBER: US/10/282,122A

75 CURRENT FILING DATE: 2003-02-20

76 PRIOR FILING DATE: 2000-03-21

77 PRIOR FILING DATE: 2000-05-23

78 PRIOR FILING DATE: 2000-05-26

79 PRIOR FILING DATE: 2000-09-06

80 PRIOR FILING DATE: 2000-09-29

81 PRIOR FILING DATE: 2000-10-23

82 PRIOR FILING DATE: 2000-11-27

83 PRIOR FILING DATE: 2000-12-22

84 PRIOR FILING DATE: 2001-02-09

85 PRIOR FILING DATE: 2001-02-16

86 Remaining Prior Application data removed - See File Wrapper or PALM.

87 NUMBER OF SEQ ID NOS: 78614

88 SOFTWARE: Patent in version 3.1

89 SEQ ID NO 59260

90 LENGTH: 352

91 TYPE: PRT

92 ORGANISM: Klebsiella pneumoniae

93 FEATURE:

94 NAME/KEY: MISC FEATURE

95 LOCATION: (209)..(209)

96 OTHER INFORMATION: X=any amino acid

97 FEATURE:

98 NAME/KEY: MISC FEATURE

99 LOCATION: (225)..(225)

100 OTHER INFORMATION: X=any amino acid

101 FEATURE:

102 NAME/KEY: MISC FEATURE

103 LOCATION: (240)..(240)

104 OTHER INFORMATION: X=any amino acid

105 FEATURE:

106 NAME/KEY: MISC FEATURE

107 LOCATION: (248)..(248)

108 OTHER INFORMATION: X=any amino acid

109 FEATURE:

110 NAME/KEY: MISC FEATURE

111 LOCATION: (296)..(296)

112 OTHER INFORMATION: X=any amino acid

113 FEATURE:

114 NAME/KEY: MISC FEATURE

LOCATION: (321)...(321)
OTHER INFORMATION: X=any amino acid
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (348)...(348)
OTHER INFORMATION: X=any amino acid
US-10-282-122A-59260

Query Match 54.8%; Score 40; DB 12; Length 352;
Best Local Similarity 66.7%; Pred. No. 1.2e+03;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 WTNLW 7
DB 266 WTKLW 271

RESULT 58

US-10-369-493-8708
Sequence 8708, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:

APPLICANT: Cao, Yongwei;
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

PRIOR FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 8708

LENGTH: 460

TYPE: PRT

ORGANISM: Ralstonia metallidurans

US-10-369-493-8708

Query Match 54.8%; Score 40; DB 15; Length 460;
Best Local Similarity 71.4%; Pred. No. 1.5e+03;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNIW 7
DB 27 EWTNDW 33

RESULT 59

US-10-369-493-16822
Sequence 16822, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:

APPLICANT: Cao, Yongwei

APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.

APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 16822

LENGTH: 591

TYPE: PRT

ORGANISM: Caulobacter crescentus

US-10-369-493-16822

Query Match 54.8%; Score 43; DB 15; Length 591;
Best Local Similarity 62.5%; Pred. No. 1.8e+03;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 WTNLW 9
DB 112 WTNLW 119

RESULT 60

US-09-779-451-74

Sequence 74, Application US/09779451

Patent No. US2002009452A1

GENERAL INFORMATION:

APPLICANT: Wild, Carl T.

APPLICANT: Allaway, Graham P.

TITLE OF INVENTION: Assay for Detection of Viral Fusion Inhibitors

FILE REFERENCE: 1900.0300003

CURRENT APPLICATION NUMBER: US/09/779,451

CURRENT FILING DATE: 2001-08-17

PRIOR APPLICATION NUMBER: US 60/235,901

PRIOR FILING DATE: 2000-09-28

PRIOR APPLICATION NUMBER: US 60/181,543

PRIOR FILING DATE: 2000-02-10

NUMBER OF SEQ ID NOS: 77

SOFTWARE: PatentIn version 3.0

SEQ ID NO 74

LENGTH: 36

TYPE: PRT

ORGANISM: Human immunodeficiency virus type 1

US-09-779-451-74

Query Match 53.4%; Score 39; DB 9; Length 36;
Best Local Similarity 62.5%; Pred. No. 2.7e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNIW 8
DB 28 EWASINW 35

RESULT 61

US-09-809-060-49

Sequence 49, Application US/09809060

Publication No. US20020010317A1

GENERAL INFORMATION:

APPLICANT: Wild, Carl T.

APPLICANT: Allaway, Graham P.

TITLE OF INVENTION: A Method for Generating Immunogens that Elicit

TITLE OF INVENTION: Neutralizing Antibodies against Fusion-Active

TITLE OF INVENTION: Regions of HIV Envelope Proteins

FILE REFERENCE: 1900.0260001

CURRENT APPLICATION NUMBER: US/09/809,060

CURRENT FILING DATE: 2001-03-16

PRIOR APPLICATION NUMBER: US 60/189,981

PRIOR FILING DATE: 2000-03-17

NUMBER OF SEQ ID NOS: 87

SOFTWARE: PatentIn version 3.0

SEQ ID NO 49

LENGTH: 36

TYPE: PRT

ORGANISM: Human immunodeficiency virus type 1

US-09-809-060-49

Query Match 53.4%; Score 39; DB 12; Length 36;
Best Local Similarity 62.5%; Pred. No. 2.7e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNIW 8
DB 28 EWASINW 35


```
RESULT 62
US-09-779-451-72
; Sequence 72, Application US/09779451
; Patent No. US20020094521A1
; GENERAL INFORMATION:
; APPLICANT: Wild, Carl T.
; APPLICANT: Allaway, Graham P.
; TITLE OF INVENTION: Assay for Detection of Viral Fusion Inhibitors
; FILE REFERENCE: 1900.0300003
; CURRENT APPLICATION NUMBER: US/09/779,451
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 60/235,901
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/181,543
; PRIOR FILING DATE: 2000-02-10
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 72
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-779-451-72

Query Match      53.4%; Score 39; DB 9; Length 46;
Best Local Similarity 62.5%; Pred. No. 3.2e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 EWTNIWNV 8
Db      38 EWASINWV 45

RESULT 63
US-09-809-060-47
; Sequence 47, Application US/09809060
; Publication No. US20020010317A1
; GENERAL INFORMATION:
; APPLICANT: Wild, Carl T.
; APPLICANT: Allaway, Graham P.
; TITLE OF INVENTION: A Method for Generating Immunogens that Elicit
; TITLE OF INVENTION: Neutralizing Antibodies against Fusion-Active
; TITLE OF INVENTION: Regions of HIV Envelope Proteins
; FILE REFERENCE: 1900.0260001
; CURRENT APPLICATION NUMBER: US/09/809,060
; CURRENT FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: US 60/189,981
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 47
; LENGTH: 46
; TYPE: PRT
; ORGANISM: Human immunodeficiency virus type 1
US-09-809-060-47

Query Match      53.4%; Score 39; DB 12; Length 46;
Best Local Similarity 62.5%; Pred. No. 3.2e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 EWTNIWNV 8
Db      38 EWASINWV 45

RESULT 64
US-10-106-698-4727
; Sequence 4727, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005F1

; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4727
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-4727

Query Match      53.4%; Score 39; DB 14; Length 47;
Best Local Similarity 44.4%; Pred. No. 3.3e+02;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      1 EWTNIWNV 9
Db      20 KFSQAWNV 28

RESULT 65
US-10-425-114-72688
; Sequence 72688, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 72688
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4726-014-F9_FLI.pep
US-10-425-114-72688

Query Match      53.4%; Score 39; DB 12; Length 55;
Best Local Similarity 42.9%; Pred. No. 3.7e+02;
Matches 3; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY      4 NIWNVAK 10
Db      47 SLWWSR 53

RESULT 66
US-09-879-957-112
; Sequence 112, Application US/09879957
; Patent No. US20020034755A1
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. US20020034755A1b
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWLES, Dana M.
; APPLICANT: MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; TITLE OF INVENTION: USING SAME
; NUMBER OF SEQUENCES: 227
```

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Pernie & Edmonds LLP
;; STREET: 1155 Avenue of the Americas
;; CITY: New York
;; STATE: New York
;; COUNTRY: USA
;; ZIP: 10036-2711
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-TOS/MS-DOS
;; SOFTWARE: Patentin Release #1.3, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/879,957
;; FILING DATE: 13-Jun-2001
;; CLASSIFICATION: <Unknown>
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/630,915
;; FILING DATE: 03-APR-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Misrock, S. Leslie
;; REGISTRATION NUMBER: 18,872
;; REFERENCE/DOCKET NUMBER: 1101-174
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 790-9090
;; TELEFAX: (212) 869-8864/9741
;; TELEX: 66141 PENNIE
;; INFORMATION FOR SEQ ID NO: 112:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 60 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: <Unknown>
;; TOPOLOGY: unknown
;; MOLECULE TYPE: peptide
;; SEQUENCE DESCRIPTION: SEQ ID NO: 112:
US-09-879-957-112

Query Match 53.4%; Score 39; DB 9; Length 60;
Best Local Similarity 80.0%; Pred. No. 4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 WWAK 10
DB 4C WWAR 44

RESULT 67
US-10-424-599-161138
; Sequence 161138, Application US/10424595
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 161138
; LENGTH: 64
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_116525C.1.pep
US-10-424-599-161138

Query Match 53.4%; Score 39; DB 12; Length 64;
Best Local Similarity 50.0%; Pred. No. 4.2e+02;
Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNIWWM 8
DB 51 KYTRSWWM 58

RESULT 68
US-10-163-587A-27
; Sequence 27, Application US/10163587A
; Publication No. US20030096263A1
; GENERAL INFORMATION:
; APPLICANT: Oliveira, Marcos
; TITLE OF INVENTION: SELECTIVE PARP-1 TARGETING FOR DESIGNING CHEMO/RADIO SENSITIZI
; FILE REFERENCE: 50229-306
; CURRENT APPLICATION NUMBER: US/10/163,587A
; CURRENT FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: 60/296,110
; PRIOR FILING DATE: 2001-06-07
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 27
; LENGTH: 71
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-163-587A-27

Query Match 53.4%; Score 39; DB 14; Length 71;
Best Local Similarity 80.0%; Pred. No. 4.5e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 WWAK 10
DB 4C WWAR 44

RESULT 69
US-09-864-408A-7046
; Sequence 7046, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shinkets, Richard A.
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides E
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7046
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-408A-7046

Query Match 53.4%; Score 39; DB 11; Length 73;
Best Local Similarity 66.7%; Pred. No. 4.6e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 5 IWWAK 10
DB 62 IWWGR 67

RESULT 70
US-10-263-103-33
; Sequence 33, Application US/10263103
; Publication No. US20030138445A1
; GENERAL INFORMATION:
; APPLICANT: AVENTIS PASTEUR
; APPLICANT: Chevalier, Michel
; APPLICANT: El Habib, Raphaelle
; APPLICANT: Krell, Tino

APPLICANT: Sodoyer, Regis
TITLE OF INVENTION: gp41 antigen
FILE REFERENCE: 01-1692-A
CURRENT APPLICATION NUMBER: US/10/263,103
CURRENT FILING DATE: 2002-10-02
NUMBER OF SEQ ID NOS: 35
SOFTWARE: PatentIn version 3.1
SEQ ID NO 33
LENGTH: 108
TYPE: PRT
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: polypeptide
US-10-263-103-33

Query Match 53.4%; Score 39; DB 14; Length 108;
Best Local Similarity 40.0%; Pred. No. 6.3e+02;
Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 EWTNIWWAX 10
: : : : :
Db 40 KWASLWVWAR 49

RESULT 71

US-09-764-891-5219
Sequence 5219, Application US/09764891
Publication No. US20030077808A1
GENERAL INFORMATION:

APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5219
LENGTH: 110
TYPE: PRT
ORGANISM: Homo sapiens
US-09-764-891-5219

Query Match 53.4%; Score 39; DB 10; Length 110;
Best Local Similarity 55.6%; Pred. No. 6.4e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 2; Gaps 1;

QY 2 WTNIW--WW 8
: : : : :
Db 77 WINMRCWW 85

RESULT 72

US-10-424-599-171353
Sequence 171353, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:

APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 171353
LENGTH: 119
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_125748C.1.pep

US-10-424-599-171353

Query Match 53.4%; Score 39; DB 12; Length 119;
Best Local Similarity 62.5%; Pred. No. 6.8e+02;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 WTNIWWWA 9
: : : : :
Db 102 WNKIGWWA 109

RESULT 73

US-09-833-245-577
Sequence 577, Application US/09833245
Publication No. US20040010134A1
GENERAL INFORMATION:

APPLICANT: Human Genome Sciences, Inc.
TITLE OF INVENTION: Albumin Fusion Proteins
FILE REFERENCE: PF546PCT
CURRENT APPLICATION NUMBER: US/09/833,245
CURRENT FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 60/229,358
PRIOR FILING DATE: 2000-04-12
PRIOR APPLICATION NUMBER: 60/256,931
PRIOR FILING DATE: 2000-12-21
PRIOR APPLICATION NUMBER: 60/199,384
PRIOR FILING DATE: 2000-04-25
NUMBER OF SEQ ID NOS: 2267
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 577
LENGTH: 127
TYPE: PRT
ORGANISM: Homo sapiens
US-09-833-245-577

Query Match 53.4%; Score 39; DB 11; Length 127;
Best Local Similarity 30.0%; Pred. No. 7.2e+02;
Matches 3; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 EWTNIWWAX 10
: : : : :
Db 31 QWGLWFWVR 40

RESULT 74

US-09-833-245-579
Sequence 579, Application US/09833245
Publication No. US20040010134A1
GENERAL INFORMATION:

APPLICANT: Human Genome Sciences, Inc.
TITLE OF INVENTION: Albumin Fusion Proteins
FILE REFERENCE: PF546PCT
CURRENT APPLICATION NUMBER: US/09/833,245
CURRENT FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: 60/229,358
PRIOR FILING DATE: 2000-04-12
PRIOR APPLICATION NUMBER: 60/256,931
PRIOR FILING DATE: 2000-12-21
PRIOR APPLICATION NUMBER: 60/199,384
NUMBER OF SEQ ID NOS: 2267
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 579
LENGTH: 127
TYPE: PRT
ORGANISM: Homo sapiens
US-09-833-245-579

Query Match 53.4%; Score 39; DB 11; Length 127;
Best Local Similarity 30.0%; Pred. No. 7.2e+02;
Matches 3; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 EWTNIWWAX 10

Db 31 OWLGLWFWVR 40

RESULT 75
US-10-424-599-281013
; Sequence 281013, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 281013
; LENGTH: 134
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(134)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_95777C.1.pep
US-10-424-599-281013

Query Match: 53.4%; Score 39; DB 12; Length 134;
Best Local Similarity 44.4%; Pred. No. 7.5e+02;
Matches 4; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
CY 2 WTNIWWAK 10
Db 66 WVGLXWWSK 74

Search completed: June 9, 2004, 18:17:01
Job time : 68 secs